Understanding Consciousness from a Physiological Perspective
A Dialogue with L. Andrew Coward

Rachel St. Clair

Interviewers: Rachel St. Clair, Susan Schneider, Ph.D., and Elan Barenholtz, Ph.D. Editing assistance by William Edward Hahn, Ph.D.

Rachel St. Clair is a doctoral candidate in FAU's Center for Complex Systems and Brain Sciences, holding appointments with FAU's Center for Future Mind, Graduate Neuroscientist Training Program and Machine Perception and Cognitive Robotics (MPCR Lab.

L. Andrew Coward was born and educated in England, graduating in natural sciences from Downing College, Cambridge. He had a 30-year career with Bell Northern Research/Nortel Networks in Canada and the United States. In the course of his career, he worked on many different aspects of the design of real time electronic control systems with billions of components. He became interested in applying the techniques for designing complex systems to understanding the brain, and his first book on this topic was published in 1990. In 1998 he took early retirement from Nortel to work full time on understanding the brain. He lives in Vancouver, Canada, but for many years has carried out research and teaching with the Australian National University where he is currently an honorary associate professor.

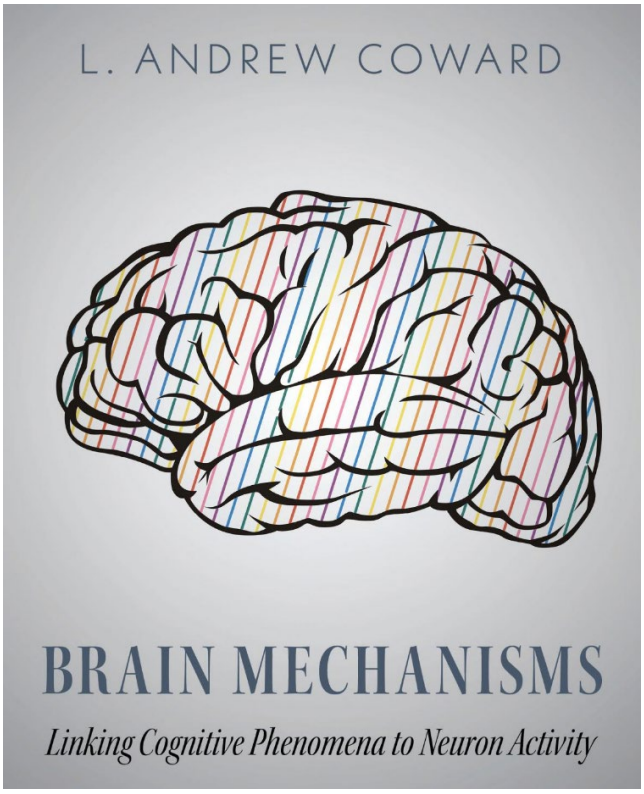
Susan Schneider, Ph.D., is the founding director of FAU's Center for Future Mind. The center explores scientific and philosophical innovations to achieve a richer understanding of emerging technologies and the future of the mind. Schneider is the William F. Dietrich distinguished professor of philosophy in the Dorothy F. Schmidt College of Arts and Letters, and a member of the FAU Stiles-Nicholson Brain Institute.

Elan Barenholtz, Ph.D., and William Hahn, Ph.D., are co-founding directors of the MPCR Lab and co-founders of the Rubin and Cindy Gruber Sandbox, an artificial intelligence lab on FAU's Boca Raton campus. The lab is a multidisciplinary research group dedicated to understanding the foundational mechanisms of intelligent systems. Since 2015, the MPCR has engaged hundreds of people across academic levels, scientific disciplines and industries in the new field of deep learning artificial intelligence and its many applications. Barenholtz is a professor of psychology, and Hahn is an assistant professor of mathematics, both at FAU.

Consider how you start your day, each morning. As you open your eyes, light photons inundate your retina – a new day of visual experience begins. Amazingly, those rapidly moving photons create tiny electrical impulses in neurons, with less voltage than a AA battery, and this leads to a cascade of brain activity. Together with other conscious activities in your brain, this gives you the feeling of being awake and alive. Conscious moments like these are an essential part of what it means to be you.

As you opened your eyes, somewhere along the way, the brain's neuronal firing transitioned from an undetected biological phenomenon to an experience that is available for the conscious control of thought and action. Consciousness is hard to define since it is a unique experience for each person (and maybe even each organism). Even so, we all understand the word in our own way. We each have some idea what it is like to be conscious; to experience the core of existence as a living being. So, how can we start to understand consciousness in terms of anatomy and physiology?

Prominent researcher, Andrew Coward from the Australian National University where he is currently an honorary associate professor, has been studying the brain fulltime since 1998. In his most recent book, *Brain Mechanisms: Linking Cognitive Phenomena to Neuron Activity*, Coward has detailed a framework for understanding conscious mental processes in terms of anatomy and physiology. In this conversation, we ask Coward some intriguing questions on the underlying physiological mechanisms of consciousness and how his framework might relate to other prominent accounts.

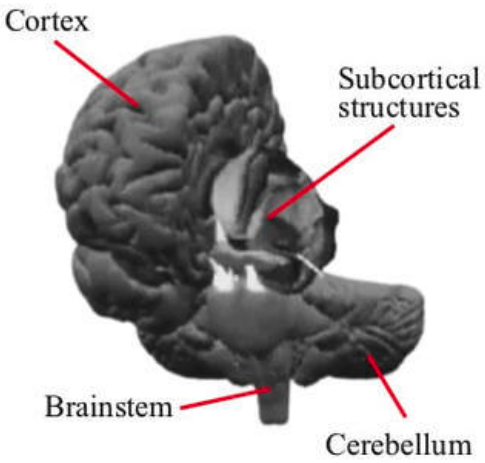


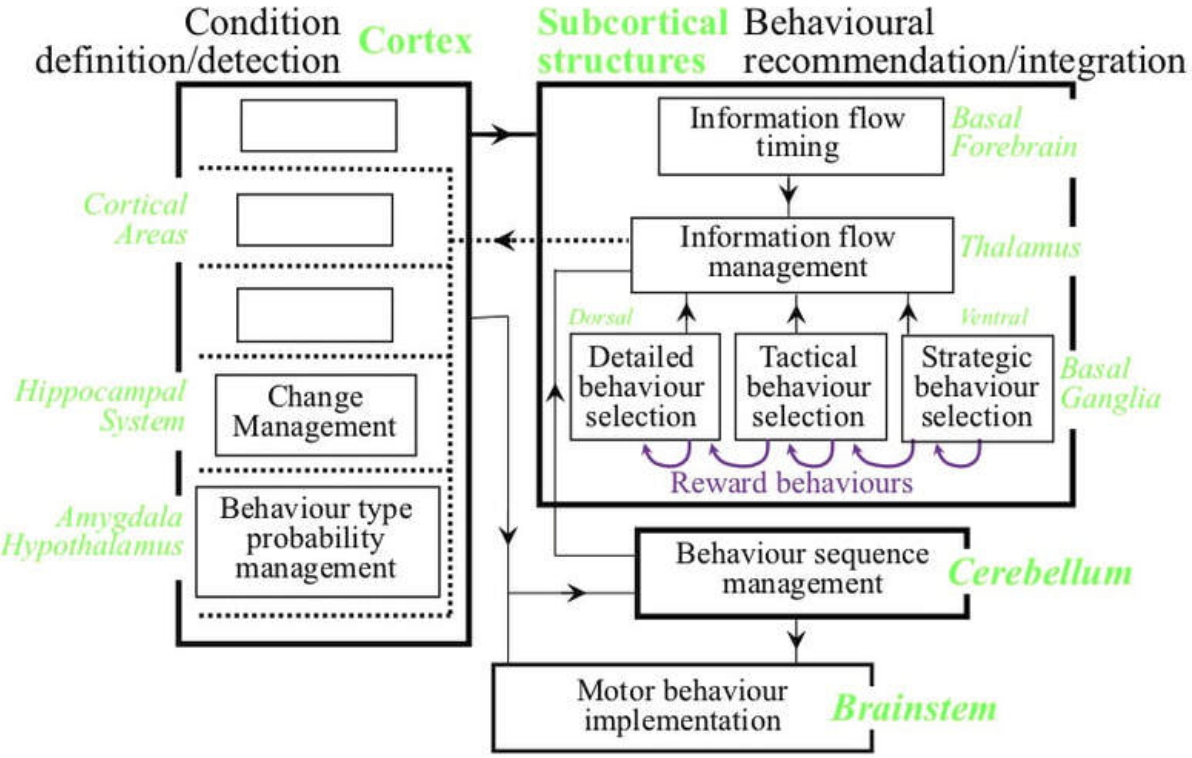
Understanding recommendation architecture (RA) requires thinking about the brain as a system in which different anatomical structures like the cortex, basal ganglia, hippocampus, thalamus, amygdala and cerebellum are subsystems that perform different, distinctive types of information processes, each contributing in key ways to the brain's conscious activity. Since the entirety of the framework is not specific to consciousness, but rather tailored to explain higher

cognitive phenomena, we will first give a quick overview and introduce some important terms in RA and then return to how RA explains consciousness.

In the human brain, signals containing information about the environment and about the state of the brain itself flow through a network of neurons, changing them so they will be more likely to fire under similar conditions in the future. This flow, or cascade of pattern extraction through multiple cortical areas, recruits the system to do something. By combining patterns of patterns, in a hierarchical fashion, the brain is able to make use of combinatorial expression of those patterns for recommending behaviors. Here, selecting the "do something" part is critically distinct from the pattern extraction hierarchy. The brain uses pattern extraction as an expression of conditions to recommend behaviors. According to RA, the pattern extraction is largely dependent on the cortex while the behavioral interpretation of those patterns is determined and implemented by subcortical structures.

In this context, the word "pattern" does not imply any exact correspondence with discrete objects in the world (e.g. Halle Berry neurons). Patterns detected by cortical neurons are simply groups of inputs that have often been active at similar times in the past. The simultaneous activity may have occurred in different cognitive situations, so it is better to refer to the circumstances in which neurons fire as information conditions (or just conditions) which are defined (i.e. stored) and later detected. This condition detection and definition (CDD) occurs largely in the cortex subsystem. This subsystem is separate from the behavioral selection (BS) subsystem, which comprises subcortical regions, namely the basal ganglia, and parts of the limbic system. The picture below outlines brain regions and their respective functions in RA.

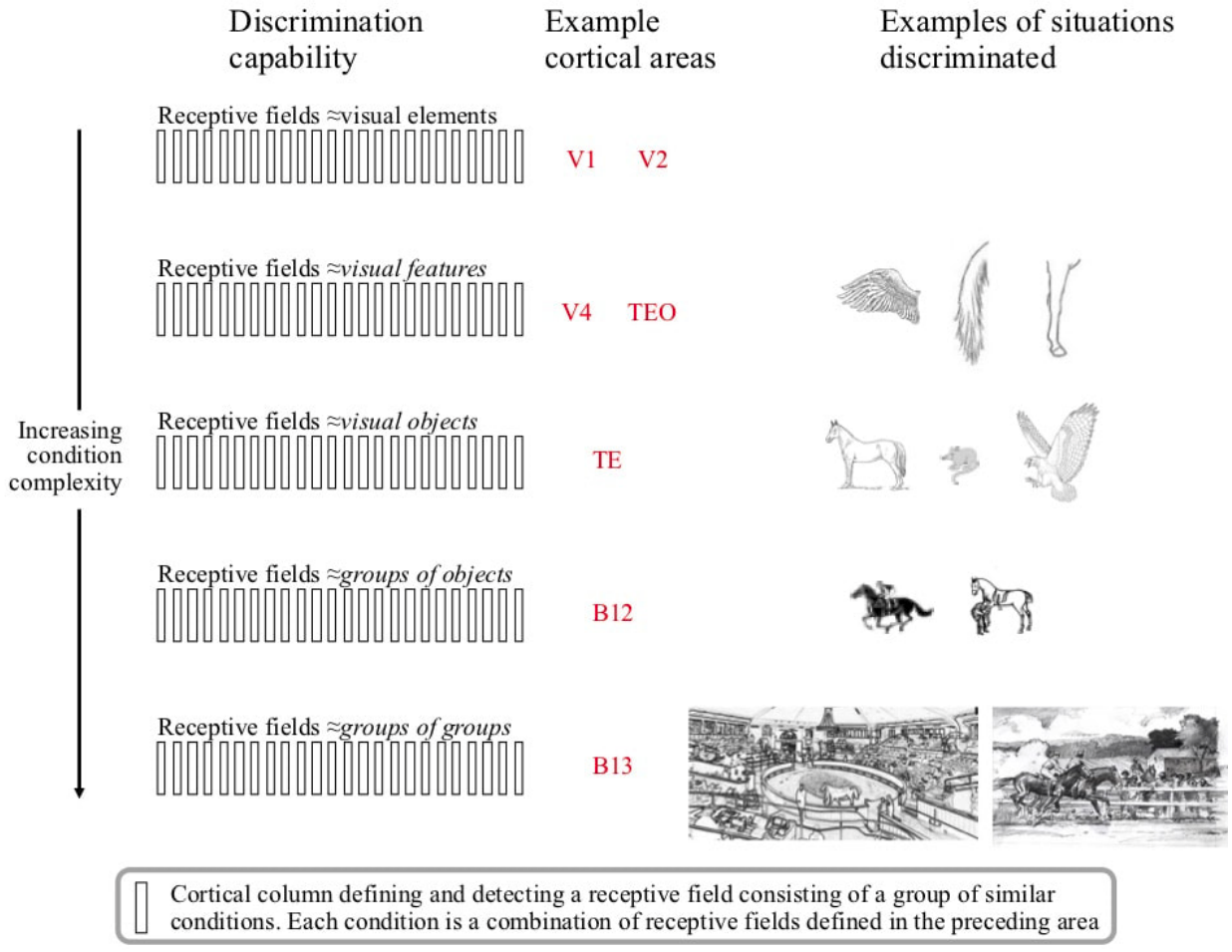




The CDD encodes information while the BS looks at those encodings as recommendations for further activity. Patterns that encode a specific condition are referred to by Coward as receptive fields, which are learned through a process of connecting inputs from other neurons and increasing connection weights. This learning process is called *receptive field expansion.* The term receptive field originally described the area of the retina in the eye from which the neuron derived all its inputs; however, Coward has generalized the term here to refer to the set of circumstances in which a brain module produces outputs.

An active receptive field provided by the CDD corresponds with neither a categorical feature nor a behavior; it is an abstract encoding of the information available to the cortex. Each active receptive field received by the BS is interpreted as a set of recommendations in favor of different behaviors; hence the name recommendation architecture. The BS determines and implements the behaviors with the largest total recommendation strength. Behaviors can be internal processes in the brain or external behaviors of the muscles.

In the RA, a neuron is directly activated by the presence of its condition within current sensory inputs. A neuron can also be indirectly activated on the basis of temporal correlations in the past activity of the neuron and groups of currently active other neurons. There are four types of correlations which can result in an indirect activation: recent simultaneous activity; frequent past simultaneous activity; past activity during a period in which a lot of receptive field expansions were occurring; and current activity (which effectively prolongs activity). These four types of indirect activation support, respectively, priming, semantic, episodic and working memory. Priming is observed when brief exposure to memory can influence our behavior, even if we are not consciously aware. Semantic memory occurs when we recall facts, while episodic memory is for events. Working memory is the ability to keep a small amount of information available for a short amount of time.



To ensure there are sufficient recommendations to support a high integrity behavior selection by the BS, at least a minimum number of receptive field detections are generated in all cortical areas through which the pattern extraction cascade passes. Receptive field expansions occur if necessary to bring the number of detections up to the minimum. Expansions are managed by a hippocampus-cortical circuit acting via cortical columns. Columns are groups of pyramidal neurons with strong vertical connectivity that increase in receptive field complexity as the information flows from layer to layer in the cortical sheet (see picture). By analogy, cortical columns can be thought of as buckets that fill and pour over to the next hierarchy of buckets, passing information along. If there is not enough water in one row of buckets, this hippocampus-cortical circuit will activate water in relevant buckets to add to the total active receptive field population, effectively recruiting contextual knowledge to the current sensory input. Note that this provides the reason for the existence of columns in the cortex - they generate the information needed by the hippocampal system to manages changes to cortical receptive fields

In summary, RA describes cognitive phenomena on multiple levels of brain activity. The framework relies on cortical columns defining and detecting input circumstances which can then be later interpreted for appropriate behavior by subcortical regions. Key ideas are summarized in the table below.

Receptive Fields	"..set of conditions defined and detected by a module (i.e. neuron, column, or cortical area)" pg.39 from *Brain Mechanisms: Linking Cognitive Phenomena to Neuron Activity*
Receptive Field Expansion	"When a condition is added to a module, the range of circumstances in which it will detect its receptive field is expanded.. Process of learning" pg. 52 from *Brain Mechanisms: Linking Cognitive Phenomena to Neuron Activity*
Direct Activation	Activate receptive fields based on current sensory input.
Indirect Activation	Activate receptive fields based on correlation of past receptive field activity. pg. 79-81 from *Brain Mechanisms: Linking Cognitive Phenomena to Neuron Activity*

Now that we have a brief understanding of what the recommendation architecture is, in what follows, we discuss various issues with Coward.

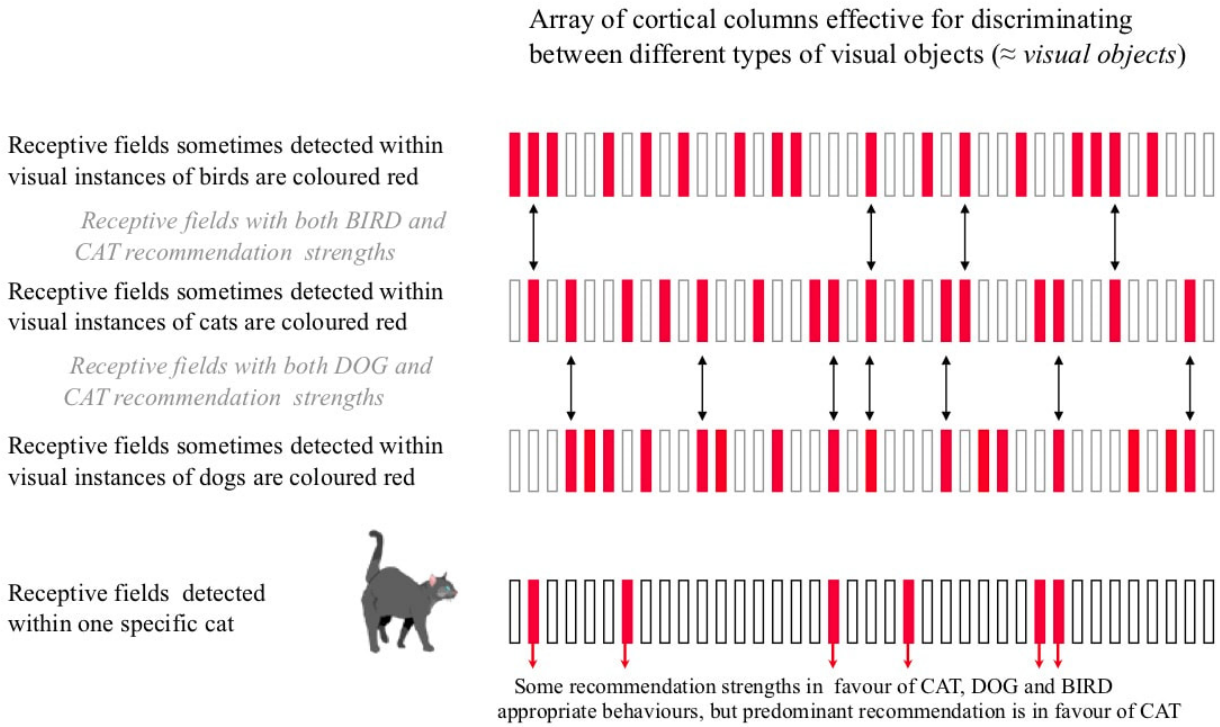
- How does recommendation architecture describe cognitive phenomena?
- How is consciousness explained by recommendation architecture?
- What areas of the brain are involved in consciousness?
- Are animals conscious?
- How does recommendation architecture relate to other explanations of consciousness?

St. Clair: Can you elaborate on what we mean by the term 'recommendation' and how recommendations impact behaviors or cognition?

Coward: In the recommendation architecture model there are three general types of behavior. One type is the release of receptive field detections. A release can be into the cortex (i.e. attention behaviors), between cortical areas (i.e. cognitive behaviors) or out of the cortex (i.e. motor behaviors). The second type is changes to recently used recommendation strengths. The third type is receptive field expansions. All three types of behavior are recommended by cortical receptive field detections. For the first two types of behavior, the basal ganglia determines which behaviors have sufficient total recommendation strength across all currently detected receptive fields to be implemented. The hippocampus determines which behaviors of the third type will be implemented.

A cognitive process will consist of a long sequence of releases between cortical areas, often involving receptive field expansions, and sometimes involving changes to recommendation weights. There may be no motor behavior, but the long-term effects of the process are the receptive field expansions and recommendation weight changes left at the end. It is these changes that instantiate gains in intelligence and understanding.

To give a simple example, suppose that a brain has often seen different cats, and at the same time the word "cat" was heard. Each time a cat was seen, a different group of receptive fields was activated, but because of the similarities between cats, some receptive fields were activated relatively often. So, a group of visual receptive fields exists that is often active at the same time when cats were seen. Similarly, a group of auditory receptive fields exists that is often active at the same time when the word "cat" is heard. Receptive fields corresponding with different subsets of that group of auditory fields can be defined, and acquire recommendation strengths in favor of indirectly activating the visual fields. Hence when the word "cat" is heard, a pseudo-visual experience of an actual cat will be experienced. The combination of receptive field and recommendation weight definitions results in a semantic memory associating the word with the visual category."



St. Clair: Can you briefly explain how these brain processes relate to consciousness? Specifically, how it might relate to what philosophers call "phenomenal consciousness" in which there is a certain felt quality of a given experience, e.g., as when we hear the sound of a familiar voice, see the vivid hues of a sunset, etc.?

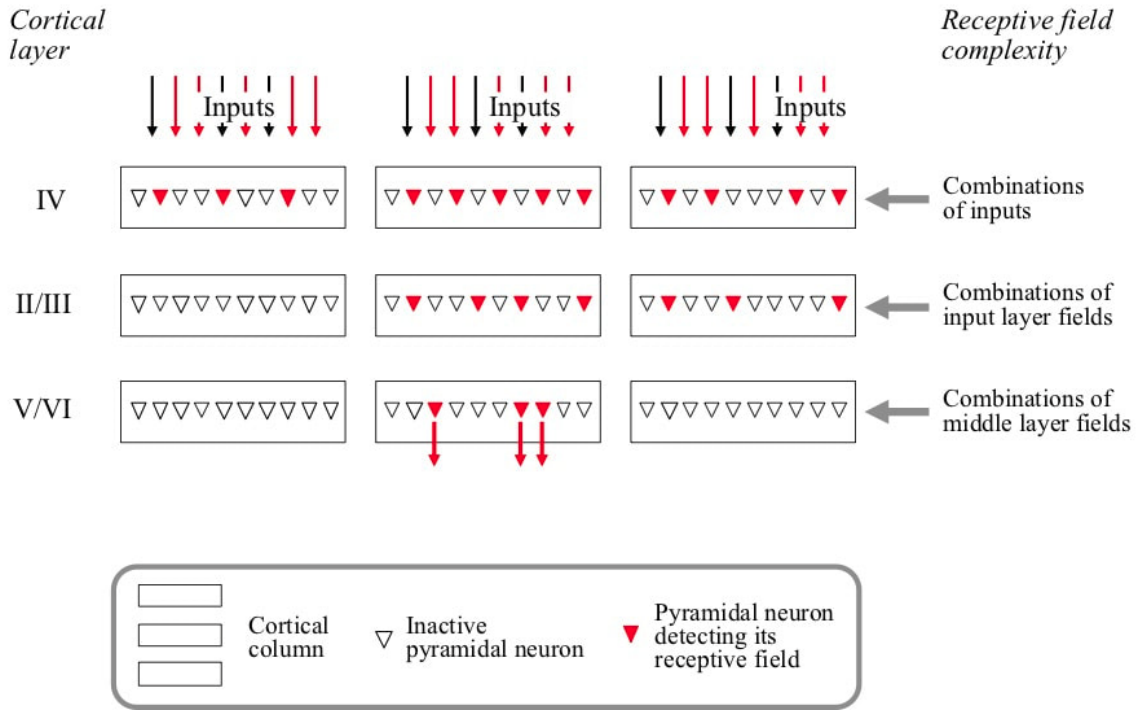
Coward: I define 'phenomenal consciousness' as the vivid experience we can sometimes get in response to a sensory input. The way I interpret this is that just looking at something (such as the color red in the common example) initially results in direct cortical receptive field detections mainly in the visual areas. Phenomenal consciousness of the color results when recommendations are accepted in favor of indirect activation of a wide range of other receptive fields active in the past at the same time as the directly detected receptive fields. The recommendations could be generated by the directly detected visual receptive fields but will probably need to be reinforced by recommendation strengths of fields detected in response to more general circumstances, which are located in higher cortical areas like the prefrontal. The fields that are indirectly activated could also be scattered across many different cortical areas. So lots of areas are involved in both the generation of the indirect activations and the content of those indirect activations. Which areas contribute most strongly will depend on the nature of the initiating sensory experience and the general circumstances of the brain at the time.

St. Clair: What do you see as the role of subcortical regions in consciousness? Would animals that don't have a cortex be “beast machines”, computing cognitive processes, but not conscious of them, merely having a sort of subconscious neuronal firing mechanism? Is there such a distinction to be made in recommendation architecture and if so, where does it fit in?

Coward: Let me start with the first part of your question: whether some brain region, process or state is involved in consciousness or not, in particular as applied to the basal ganglia.

I am a little cautious about this question. I can illustrate the reason for my caution with a computer example. In a computer there are different major hardware components like the central processing unit (CPU), random access memory, WiFi interface, screen interface, keyboard interface etc. ... Each of these components has subcomponents like integrated circuits, and each of the integrated circuits has substructures and sub-substructures all the way down to transistors. If you asked the question which of the major components is involved in an application like web browsing, the answer would be all of them. In fact, all these components are involved in almost every application. Even at a very detailed level, many of the individual transistors could be involved in every application (one especially clear example is the transistors implementing the basic instruction set of the CPU). Essentially, components on every level specialize in different types of information processes, and any one type is typically required to some degree by every application.

I see an analogous situation in the case of the brain, with different brain structures specializing in different types of information processes (although all the information processes are completely different from those in computers). Hence, I prefer to ask the question: what are the information processes performed by different brain structures in support of consciousness? Of course, some processes are more complex, and the question could be rephrased as: what brain regions perform the most complex information processes in support of consciousness?



At any point in time, I see the *cortical activity* supporting consciousness as a population of *pyramidal neurons* generating outputs indicating detections of their receptive fields.[1] Each layer V pyramidal targets *medium spiny neurons (MSNs) in the basal ganglia*. An MSN corresponds with a behavior where behaviors are often releases of information.[2] In many cases these releases are from one cortical area to another, in some cases releases are out of the cortex to drive motor behaviors or behaviors of focusing attention on specific subsets of sensory inputs. In information terms, an active output from a pyramidal neuron on to an MSN is a recommendation in favor of the behavior corresponding with the MSN, and the weight of the synapse making the connection is the weight of the recommendation in favor of the behavior. The basal ganglia determines the total recommendation weights of all behaviors across all currently active pyramidal neurons, and implements the most strongly recommended behaviors. A release is implemented by the basal ganglia triggering a part of the thalamus to impose a frequency modulation (or resonance) on some cortical outputs. Such a release drives

[1] Cortical layers are highly-laterally connected areas of neurons which allow for receptive fields to be expressed in a combinatorial fashion (i.e that patterns of patterns can be detected). Sensory input comes into layer IV and moves through layers II/III as receptive field detections elicit a cascade of neuronal firing, until it reaches layers V/VI. Layer V provides outputs to the basal ganglia, and layer VI outputs go elsewhere in the cortex.

[2] Behaviors can be internal processes or external muscle commands.

activity in the targeted cortical area. So the basal ganglia determines (on the basis of its cortical inputs) which other neuronal regions will be recruited to contribute to overall cortical activity.

There's a lot of argument about the exact definitions of conscious and subconscious. I prefer to focus on understanding specific phenomena at the anatomical and neural level. The phenomena I find most relevant in this area are: the observation that we can have less vivid and more vivid experiences in response to the same stimulus (e.g. the color red) where the more vivid experiences are sometimes labelled "qualia," but are very difficult to describe verbally in detail; the experience of a constant stream of mental images, verbal and visual, with little connection with current sensory inputs; the observation that we can process very complex problems without being aware of the processing and end with a solution to the problem that can be described verbally; the observation that every one of us can be aware of a person "inside" of us that seems distinct but feels like our inner self (i.e. self awareness); and so on. I discuss all these examples in my book, *Brain Mechanisms: Linking Cognitive Phenomena to Neuron Activity*, but I will just comment on the first.

If we look at something red, our brains detect a range of receptive fields. These fields have all sorts of recommendation strengths, such as saying that is red. This is the less vivid experience. However, all the directly detected receptive fields also have recommendation strengths in favor of indirect activation of other receptive fields on the basis of temporally correlated past activity.

If these recommendations are accepted, there will be activation of a much larger population of receptive fields, resulting in a more vivid experience. One way of interpreting this larger active population cognitively is that it is made up of fragments of many different experiences that occurred when the color red was present. However, they are only fragments, and no fragment will have sufficient recommendation strength in favor of, for example, a verbal description of the past experience. So, although the experience is vivid because of the number of receptive fields activated, it is not possible to verbally describe the experience in detail, beyond saying it is more vivid.

It is possible that one fragment could have slightly larger recommendation strength in favor of further indirect activation that any other fragment, and if that recommendation were accepted it could lead to a recall of that one experience. The behavioral value of the more vivid experience is that it greatly expands the range of recommendations beyond those available in the directly detected receptive fields.

This description implicitly involves processes in both the cortex and subcortical structures. In my view, any cognitive process will require information processes performed by all the major brain structures.

All vertebrate animals have a cortex or functional equivalent (for discussion see here). All animals also have a basal ganglia and hippocampus. So, something analogous with the vivid experience could occur in other animals. I believe that maintaining useful meanings in long streams of consciousness depends on mechanisms derived from speech, and in non-human animals the lack of speech capabilities mean that streams of consciousness are much shorter and less well defined.

Schneider: Certain nonhuman animals can formulate complex plans as well as have a system of communication, however. Indeed, some cognitive scientists, such as the late Jerry Fodor, argued that nonhuman animals had an internal language of thought. Would that suffice to provide more sophisticated streams?

Coward: The internal language of thought hypothesis suggests that in the thinking process, simple concepts are combined to create more complex concepts in ways that resemble the rules of language grammar.

It is certainly true that there are structured neuron processes that support thinking, and these processes have some general resemblance to the processes that support language. Think about the way an appropriate response is generated to immediate sensory inputs. Suppose we experience a situation in which there are a number of different objects, and these objects form different groups and groups of groups. An example might be that we are walking on a city street, and there are different shops with window displays, different groups of people on the sidewalk including perhaps a couple walking a dog, cars on the road and so on. How do we think about this situation to determine an appropriate response? Our attention scans around, fixing on different objects. The process in the brain is that conditions (i.e. pyramidal neuron receptive fields) are detected in some cortical areas within several objects in succession, with the condition detections maintained active (i.e. prolonged). Then all the conditions detected in the several objects are released together to drive detections of more complex conditions in different cortical areas – these more complex conditions combine information from all the objects in a group. The more complex conditions detected within several groups are prolonged, then released to yet other cortical areas where even more complex conditions are detected that combine information from several groups. The even more complex conditions detected within several groups of groups are prolonged, then released to cortical areas where yet more complex conditions are detected, that combine information from several groups of groups.

Each detection in every area has a range of recommendation strengths in favor of different behaviors. The behaviors with the largest total recommendation strengths are carried out. So, behavior is determined by sequences of activity releases between cortical areas (and incidentally, prolongations are also implemented by releases back and forth between the prolonged areas and frontal areas). These sequences could be labelled a syntax of thought.

In this example, all the receptive fields were present in current sensory inputs. Sometimes the most appropriate behavior is not adequately recommended by the currently detected receptive fields. Receptive fields can also be indirectly activated. As mentioned earlier, a pyramidal neuron can be indirectly activated on the basis of various types of past correlations between the activity of the neuron and groups of already active neurons. Note that an indirect activation is also a behavior that must be recommended with sufficient total strength across already active pyramidal neurons. As mentioned earlier, there are four types of temporal correlation that can be used to drive indirect activations. The way in which sequences of indirect activations are managed could be regarded as part of the syntax of thought.

Language uses a relatively small and very structured information space to manage access to a much larger information space. In human beings, visual information processing is much more complex than auditory. The optic nerve carrying information from the eyes to the brain has about a million axons, while the cochlear nerve carrying auditory information from the ears only has around 30 thousand axons. The monomodal cortical areas processing visual inputs are much more extensive than the monomodal areas processing auditory inputs. Hence speech allows information from a smaller information space to access and manipulate information in a much larger information space. If one particular word is often heard at the same time as an object of a particular category is seen, the auditory receptive fields can acquire recommendation strengths in favor of indirect activation of the visual receptive fields. For example, hearing the word dog can indirectly activate the visual receptive fields often active when dogs were seen in the past. Hence the wide range of recommendation strengths associated with the visual fields is made available. Hearing the word “chase” can indirectly activate the more complex “groups of objects” visual receptive fields often active when scenes of something chasing something were seen in the past. The visual information indirectly activated by a sequence of words is processed by a sequence of activity releases between cortical areas in a way that is generally similar to the way information derived from direct visual experience is processed.

In non-human animals, the sequences of releases between cortical areas to process sensory inputs and all the mechanisms for indirect activations are generally similar to those in humans. So, in that sense you could say that even without language there is a syntax to thought.

However, as discussed earlier, if there is a long sequence of indirect activations there is a tendency for cognitive meaning to disappear. In humans, meaning is preserved by the use of the very strong indirect activation strengths created by speech capabilities. So, although chains of indirect activation are certainly possible in non-human animals, only relatively short chains will be behaviorally useful.

Tool making is an example of this. Tool making has been observed in a number of animal species, including chimpanzees, crows and dolphins as well as humans. Tool making involves behaviors to make some object that will assist other behaviors. To make a tool, receptive fields must be activated that have appropriate recommendation strengths. For example, to drive the crow muscle movements that bend a wire to make a hook for retrieving food, or the chimpanzee muscle movements to make the tip of a stick into a fan for capturing termites, or the human muscle movements to make a vast range of possible tools. Some receptive fields need to acquire recommendation strengths in favor of the muscle movements that make the tool. It is easiest for the receptive fields directly detected in the circumstances in which the tool is used to acquire the necessary recommendation strengths. If a tool is made well in advance, somehow those receptive fields must be indirectly activated in the absence of the sensory inputs that define them. Language capabilities give humans the ability to perform long chains of indirect activations, and humans can therefore make tools to work in environments completely different from their current sensory environment, even environments they have never actually experienced (such as tools to work on the surface of Mars). Crows tend to make a tool only where the tool will be used, although they can store the tool afterwards for later use. Chimpanzees also tend to make tools only in the location where they will be used, although they can collect the material needed to make the tool in advance and bring it to the location. In other words, there is some capability to support very short chains of indirect activation, but only human brains can support very long chains.

So, although the way in which sequences of activations of cortical receptive fields are organized to support determination of the most appropriate behaviors could be labelled a language of thought, this does not avoid the limitations on long chains of activations in the absence of human-like language.

Barenholtz : So according to RA, both cortical and subcortical neuronal populations are involved, to some degree, in conscious experiences. You also propose that there is no definitive line between what is conscious and what isn't. What about phenomena like blindsight and continuous flash suppression where there seems to be fairly high-level cognitive processing taking place that most researchers describe as being 'unconscious' or outside of awareness?

Coward: Both conscious and almost all unconscious processing involves receptive field detections in the cortex and determination of predominant recommendation strengths in the subcortical structures, so in that sense the highest level of physiological processes are similar. I mentioned earlier that I prefer to focus on understanding specific phenomena where differences can be identified at a more detailed level.

One often used rough indicator of whether a cognitive process is conscious or unconscious is whether it can be described verbally as it is going on and/or afterwards. I certainly agree that with this definition, very complex cognitive processing can occur in the absence of consciousness. One of the classical examples is the way in which the mathematician Poincare, with no prior thinking he was aware of, came up with the conclusion that the transformations he had used to define Fuchsian functions were identical with those of non-Euclidean geometry. Later he was able to demonstrate the truth of this conclusion by conscious reasoning. Probably, the receptive fields used for originally reaching this conclusion had very little speech recommendation strength but acquired such strength during the conscious reasoning process.

If conscious processing is defined as being capable of verbal description, an unconscious process is one that is carried out in such a way that during the process the active receptive fields do not have any predominant current recommendation strengths in favor of speech behaviors. This could occur in numbers of ways in addition to the absence of recommendation strengths in the Poincare example.

In the case of continuous flash suppression, a rapidly changing sequence of images is presented to one eye, which suppresses perception of a salient but static image presented to the other eye. In this case the recommendation strengths in favor of releasing the inputs from the dynamic images into cortical areas beyond V1 are much higher, and the inputs from the static image are not released beyond V1 (or even into V1). The receptive fields in the higher visual areas are the main contributors to receptive fields with speech recommendation strengths, and there are therefore no such strengths in favor of describing the static image. Furthermore, any ability to recall the static image is dependent on receptive field expansions in the higher visual areas. Since the input information does not reach those areas there is no receptive field expansion, and no memory.

In the case of blindsight, the behaviors in response to retinal information are managed without cortical receptive fields, and there is therefore no recommendation strength in favor of speech behaviors.

Barenholtz: So, if 1) complex processing does not necessarily entail consciousness as indicated by your Poincare example, and 2) reportability is not necessary for consciousness, as indicated by your position that animals can have consciousness in the absence of speech, then what are the criteria for calling something conscious or unconscious?

Coward: I think both the words conscious and unconscious are often used to describe many different cognitive phenomena that are supported by different detailed neuron mechanisms. What I try to do is understand how each specific phenomena can be understood in terms of neuron activity.

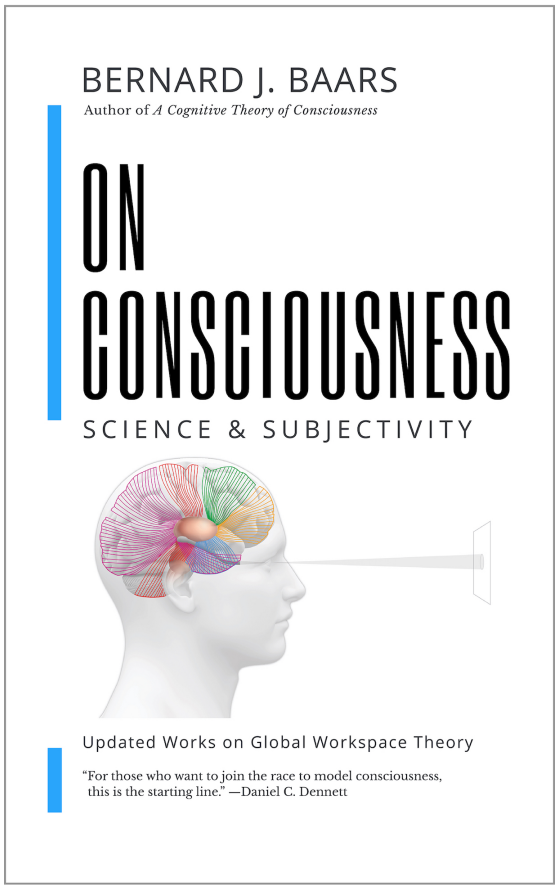
Ned Bloch proposed that there are two types of consciousness: phenomenal and access. Phenomenal consciousness is the subjective experience, which I earlier suggested is the result of a substantial indirect activation of neurons that is impossible to describe verbally. This phenomenal consciousness is certainly possible in animals. Access consciousness is a mental experience that can be described verbally, which includes the stream of consciousness. The Poincare example is neither phenomenal nor access consciousness

One comment on access type consciousness in this context is that even when we provide a verbal description of our mental state, that description will only be approximate and may even be a misleading description of why, for example, we perform some externally directed behavior. The reason for this is that our externally directed behavior is the behavior with the predominant recommendation strength across currently active receptive fields. These currently active receptive fields also have recommendation strengths in favor of speech. However, some of the receptive fields strongly recommending external behavior may have very little recommendation strength in favor of speech. To give an example, it has been observed that when a person with skill in operating some machine is asked to describe their skill, their words often reflect how they were first instructed in how to operate the machine, rather than their actual skilled operation. I think that in general the verbal reasons we give for our behavior, such as making some significant decision, are not necessarily accurate, because in an active population of receptive fields the fields with recommendation strength in favor of the behavior may not have the predominant speech recommendation strengths.

Barenholtz: It seems like your account addresses the conditions under which conscious experience occurs but, at least in my reading, not the why of inner subjective/phenomenal experience itself, what has been called the “hard problem” of consciousness. Do you think that there is anything left to be explained if we can account for observable, behavioral or physiological phenomena? If so, does your account provide some insight into the subjective nature of conscious experience?

Coward: Conscious experience in the sense of phenomenal consciousness and the stream of consciousness correspond at a more detailed level with the indirect activation of cortical and basal ganglia neurons. All these neurons are also directly activated during direct sensory experience. The major difference is that the pyramidal neurons in the primary sensory cortices that receive the first inputs from the senses are not indirectly activated, so the experiences are not visual or auditory hallucinations. With this exception, qualitatively the internal conscious experiences are very similar to direct sensory experiences, and at the neuron level both types of experience can be scientifically described in terms of the activation of populations of cortical pyramidal neurons and their associated recommendation strengths in the basal ganglia. In my view such a description is a scientific description of consciousness. One way of looking at the "hard problem" is that it asks the question of why the activation of these populations feels as it does and not feel something different. I view this as a non-scientific (perhaps philosophical) question, of the same type as the ultimate philosophical question of why something exists rather than nothing. Although such questions feel profound, I do not see any scientific approach that could make progress towards answers.

St. Clair: Now that we have a better understanding of under what conditions something is conscious and the physiological mechanisms driving that experience, how does the Recommendation Architecture explanation compare to global workspace theory (GWT), as in Bernie Baars' book, *On Consciousness*?



Coward: I would like to focus on one consciousness phenomenon and discuss how to describe that phenomenon in the two models, and where the descriptions map into each other. The phenomenon is the experience of the stream of consciousness. This experience seems to be a sequence of fairly sharp visual and/or verbal images, separated by periods in which the images are somewhat more fuzzy. We could verbally describe the content of an image at the time, and even in some cases remember and describe it later.

To summarize how I understand the GWT description, there is a workspace that contains the current image in the stream of consciousness. That content is broadcast to a large number of unconscious cognitive brain processes (i.e. receiving processes). Some of these receiving processes control voluntary actions like motor movements. Coalitions of other unconscious processes can act as input processes to the global workspace and compete for such access. Sensory inputs are briefly stored (for a few hundred milliseconds), and while stored they also compete for access to the workspace. The current workspace contents remain for a few seconds and are then replaced by new contents on the basis of the winning unconscious processes. The current contents can be described verbally.

In the RA, the current activity relevant to the contents of consciousness is all the cortical receptive field detections by layer V pyramidal neurons, as a result of direct or indirect activations. Each layer V pyramidal neuron targets a number of MSNs in the basal ganglia, and the synaptic weight of the connection is the recommendation weight of the pyramidal in favor of the behavior driven by the MSN.

The speech behavior with the largest total recommendation strength across the active layer V population is the way we would describe the current contents of our consciousness, but this is only a rough approximation of the actual contents, because many receptive fields would have other speech recommendation strengths, and some total speech recommendation strengths could be quite large, but less than the largest total.

There are many other recommendation strengths possessed by the active neuron receptive fields. Some of the key strengths supporting stream of consciousness are in favor of indirect activation of other neurons on the basis of past activity at the same time as the currently active neurons. As described earlier, the relevant types of past activity are recent simultaneous activity; frequent past simultaneous activity; past activity during a period in which a lot of receptive field expansions were occurring; and current activity (which effectively prolongs activity).

The largest total recommendation strengths of these types across the current population drive indirect activations that generate the next population of receptive field detections.

The functional value of the indirect activations is that ranges of recommendation strengths are made available that are not generated by receptive fields directly detected in current sensory inputs, but may be relevant to the current situation based on correlations in past activity. The problem is that a long chain of indirect activations risks ending up with a chaotic and cognitively meaningless spectrum of behavioral recommendations. Hence in the RA model, after a few steps of indirect activation, the predominant speech recommendation of the current population is used to construct the next population. This new population will therefore be focused on a clearer overall cognitive meaning and is then evolved by a few indirect activation steps before being focused again. This accounts for the experience of relatively sharp mental images separated by periods of more fuzziness.

The current population of receptive field detections also has recommendation strengths in favor of attention behaviors that release subsets of current sensory inputs to contribute to the next receptive field detections, and recommendation strengths in favor of a range of different

motor movements. If the total strength in favor of one of these behaviors was large enough, it would be carried out.

So, if I try to map between the two models:

The current contents of consciousness in GWT corresponds in RA with the current population of active layer V pyramidal neurons across the cortex.

Broadcast to unconscious processes in GWT corresponds in RA with the outputs from layer V pyramidal neurons generating behavioral recommendations in the basal ganglia.

Coalitions of unconscious processes can act as input processes to the global workspace, and compete for such access in GWT corresponds in RA with MSNs in the basal ganglia driving indirect activations on the basis of combinations of temporal correlations in past activity.

Sensory inputs are briefly stored (for a few hundred milliseconds), and while stored they also compete for access to the workspace in GWT corresponds in RA with prolongation (a behavior selected by the basal ganglia) of receptive fields detected within a subset of current sensory inputs. With enough recommendation strength across the population of active receptive fields, some of these sensory inputs will be released into the cortex to influence receptive field detections in the next population.

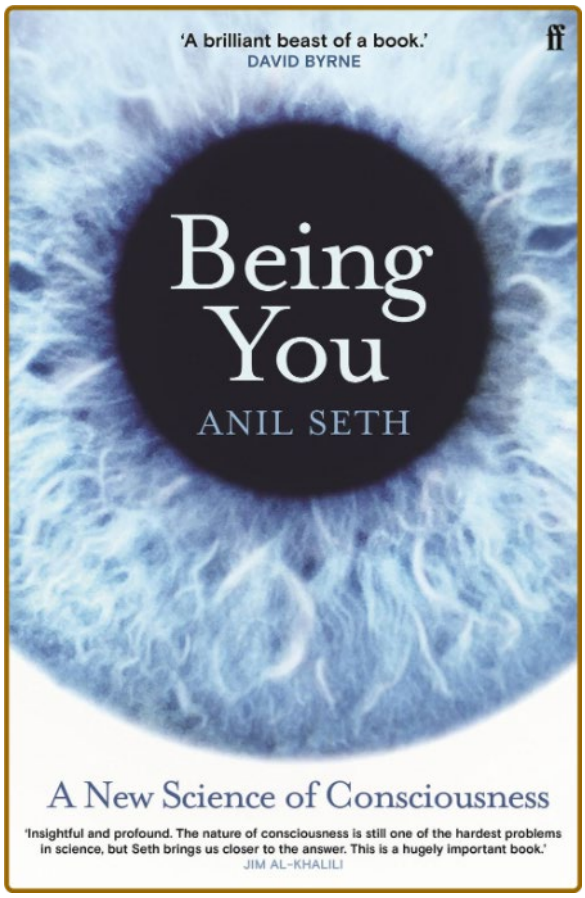
The current contents can be described verbally in GWT corresponds in RA with the predominant recommendation strengths of the active receptive fields in favor of speech behaviors. However, as discussed earlier, there are a lot of ways of understanding the content that cannot be described verbally because the total recommendation strengths are not predominant.

Some of these receiving processes control voluntary actions like motor movements in GWT corresponds in RA with some receptive fields having recommendation strengths in favor of motor movements. One subtlety here is that if these receptive fields do not have recommendation strengths in favor of verbal descriptions, the movement behavior will appear to be unconscious.

The current workspace contents remain for a few seconds and are then replaced in GWT corresponds in RA with a cognitively fairly meaningful population being replaced by another cognitively fairly meaningful population by means of a sequence of indirect activations followed by a focusing step using speech recommendation strengths.

So, the RA provides a more detailed physiological description. The key things in RA that are missing from the GWT is how the unconscious processes change the contents of consciousness, and how cognitive meaning is preserved as the contents evolve. In the RA, contents change by indirect activations of cortical pyramidal neurons on the basis of past temporal correlations in activity, and meaning is preserved by periodically using the predominant speech recommendation of the current contents to define the next current contents."

St. Clair: And how might RA compare to predictive processing, AKA predictive coding, a view held by Andy Clark, Anil Seth and others as it described Anil Seth's *Being You*?



Coward: I think that the predictive coding approach[3] is fairly easy to relate to the recommendation architecture, although in my view the use of the word prediction is anthropomorphizing neurons.

Pyramidal neurons have recommendation strengths in the basal ganglia, and these recommendation strengths are available when the pyramidal neurons are active. However, the recommendation strengths of just the pyramidal neurons that detect their receptive fields in current sensory inputs may often be inadequate to determine the most appropriate current behavior.

Pyramidal neurons don't 'know' anything about cognition or behaviors or predictions, the only thing they can 'know' is whether they are active and whether other neurons are active at the same time. That 'knowledge' can be used to broaden the range of available recommendations, by activating other neurons on the basis that they were active in the past at the same time as the currently active neurons. Once again, I identify four types of temporal correlation: frequent past simultaneous activity; recent simultaneous activity (on one occasion); past simultaneous activity at a time when receptive fields were changing; and current simultaneous activity (i.e. activity prolongation). Also, activity just before or just after the activity of a neuron can also be a useful temporal correlation. So, a population of active pyramidal neurons can indirectly activate other pyramidals on the basis of temporal correlations in past activity. The indirectly activated neurons are a kind of pseudo-sensory experience. The recommendation strengths of the indirectly activated neurons are therefore made available.

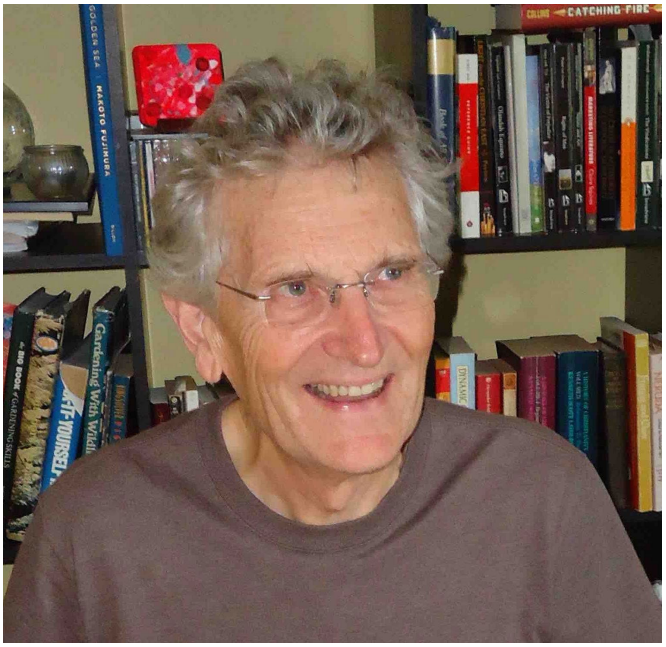
You could label such indirect activations predictions. In other words, based on current activity the brain is generating pseudo-sensory activity on the basis of past temporal correlations, and this pseudo-sensory activity is a kind of prediction of the most appropriate behavior on the basis of similarities with past experience.

To give an example, suppose you have looked at many different objects, rotating them so that you can see them from different angles. Any one receptive field will generally be detected in response to many different objects. Because you have viewed many objects in quick succession, some receptive fields have often been detected shortly after other receptive fields. This makes it possible to indirectly activate receptive fields that were often active in the past shortly after other receptive fields. If a brain was asked to imagine how a currently perceived object would

[3] Predictive Processing approaches can be summarized as an understanding of the brain where forward neuron signaling is used to communicate errors of prediction to other neurons in higher layers. The prediction is a backward neuron signal that is a guessing what the next input signal the neuron will receive. Thus, in predictive processing, the brain is constantly trying to guess what will happen next.

look if rotated, the process it could use is indirect activation on the basis of frequent past activity shortly after the currently active receptive fields. The result would be a pseudo-image of the rotated object which was an average of past experience with many other objects, weighted by the slight degree of similarity each has to the current object.

You could call this a prediction of how the object will appear when rotated. However, I personally think this is a misleading way to think about the activity, because the neurons are not really predicting anything, they are just getting activated on the basis of similarities between past and present experience.



St. Clair: Thank you Andrew for allowing us to pick your brain on the intricate processes regarding consciousness! What would you say to anyone who wants to enter the field of consciousness research?

Coward: I think many workers in the field of consciousness find it difficult to engage with all the detailed anatomy and physiology of the brain. Making such engagement possible is a key property of the recommendation architecture approach. To engage effectively, it is important to define the phenomena you are attempting to understand as specifically as possible. That is what I try to do when I discuss consciousness in my book.

How remarkable! *In summary*, we can now say that the recommendation architecture is a framework for understanding conscious and unconscious cognitive phenomena in terms of brain activity. There are five major neuronal processes that provide the understanding: condition detection and definition in the cortex, behavioral selection in subcortical areas, direct and indirect activation of receptive fields, and receptive field expansion. These physiological processes describe a scale of consciousness. Perceptions can become more or less conscious depending on how much indirect activation occurs. Reportability of cognitive events depends on multiple iterations of indirectly activated speech. Multiple cortical and subcortical brain regions are used in any one cognitive process. RA offers a unique view when compared to other approaches to understanding consciousness by explaining the phenomena in terms of physiology. Yet, there is much more discussion to be had in integrating a complete understanding of the brain and consciousness. Stay tuned for future conversations on this work and others like it.