Consciousness from an Evolutionary Point of View

Gaurav Shah

February 3, 2018

Contents

1 The Skeleton Summary

There are many different threads in this account of consciousness, and they can be woven together in many different ways. Here is one simple attempt.

1. Consciousness arises from evolution, and therefore is an adaptation that has an explanation in terms of evolutionary fitness.

2. There is nothing that conscious beings can do that non-conscious ones ("zombies") cannot. The evolutionary advantage is therefore in terms of doing the same things, but more efficiently.

3. The brain uses 20% of the calories of the body; being more efficient would therefore certainly be helpful.

4. Evolution will optimize computational ability (proxy: the weighted average of the VC dimension, and the importance of each solution) with respect to a cost function (which is the energy spent in computation, possibly combined with the number of brain cells and/or the genetic information required to specify brain development).

5. It is plausible under this constrained optimization, the neural network acquires a modular form. In particular it is more efficient if the higher-order logic can be isolated into one module, and this module can be used for many different applications (the most important at any given time). Perhaps this behavior can be compared to the self-organization

that happens in other systems when we optimize for maximum entropy flow.

6. This modularization is *not* useful unless we also have an overall module that sits on top and is able to evaluate which application gets to use the higher-order logic module at any given time. This module should be able to have access to all information in order to make this decision (introspection/qualia).

7. Understanding and dealing with other humans, who are also trying to understand us, is one of the most important problems humans evolved to solve. Introspection is an important advantage in this. As humans grow more intelligent, this becomes a more and more difficult problem to model.

8. As we are cooperative animals, the marginal advantage of intelligence keeps *increasing* as intelligence increases, in contrast to most other traits. This leads to runaway increase in intelligence.

9. As the system (in particular, the higher-order module or the control module) pushes towards higher intelligence, and approaches a critical level of complexity (some extension of Turing completeness or recursive logic), a phase transition happens and intelligence (as measured by the ability to solve a wide range of problems) increases extremely fast.

10. The fact that neural networks are highly connected means that they are much less time-reversible than computers (the state is much more complex). Also, the fact that they contain some extension of Turing completeness/recursiveness means that they are less predictable (highly connected, plus recursive looping). In connection with introspection, this leads to free will.

2 Consciousness arises from evolution

Consciousness is a product of evolution, and therefore has an explanation in terms of evolutionary fitness. All discussions of consciousness from arising from other sources, such as computers, are hypothetical: every conscious entities that we are actually aware of acquired consciousness through evolution.

Note that the brain uses up approximately 20% of the calorie expenditure of the human body, and requires better nutrition (more meat, bone marrow, cooking of food), etc. We've given up a lot for our brain capacity. Anything that makes the brain more efficient would be a great evolutionary advantage!

Of that 20%, two thirds is for actual electrical signalling involved in thinking. The other one thirds is cell maintenance, which depends on the number of cells rather than brain activity.

2.1 What is evolution optimizing for?

- Objectives:
  - The ability to solve a large range of problems, especially in situations where we are faced with multiple problems simultaneously.
  - The ability to model, and thereby correctly interact with, other humans, who are also modeling us on their part.
- Costs:
  - Energy spent on computation, which is also similar to the informational throughput.
  - Brain complexity and brain size
  - The amount of genetic code information that is required to specify brain development, perhaps.

To elaborate, the objective function is the **weighted average** of our ability to solve a wide range of problems, weighted with the importance of each of these problems (which of course, can vary with time – a saber-toothed tiger attacking *should* concentrate the mind wonderfully.

2.2 What can evolution give us that pure computation cannot?

Nothing, and everything.

As Julian Jaynes writes, there is nothing that a conscious entity can do that an unconscious one (a zombie) cannot. The advantage of evolution has to be in the ability to do **everything** more efficiently, not one particular aspect. As mentioned, this should have significant evolutionary advantages.

2.3 Some quick definitions

Intelligence is the ability to solve problems, often ones that we have not encountered before. It is an **external** characteristic, and can be measured by other observers.

Consciousness is a quale. It's an **internal** description – it's perfectly possible to have zombies that imitate conscious entities perfectly, as seen by an external observer. We do not have a definition of it, but we don't necessarily need one in order to make progress. We can go back and forth from the two axioms:

- Consciousness is an evolutionarily derived property that humans, and to a lesser extent, some animals, have. We start off with an intuitive, non-technical idea of what it is.

- Consciousness derives from the neuronal circuitry of the brain

Explaining consciousness will help us define it better. In turn, defining it better will help us pin down our explanation for it.

3 How consciousness can improve evolutionary fitness

3.1 Re-use of neural circuitry

Consciousness is associated with the existence of a higher-order processing system. This single higher-order processing module can be used for multiple applications – language, cause-and-effect reasoning, perhaps prediction of what other humans are about to do. If that were not the case, we would have completely separate circuitry for each of these, each with their separate own higher-level processing circuits.

This **in itself** does not lead to consciousness – only to more efficient neural networks.

It is true that there are definite differences between the brain structure and neural networks – however, at this early stage of understanding, we will ignore them.

3.2 Temporal Switching

Having a single higher-order processing module means we can only deal with one issue at a time. So it would be useless without the ability to choose to focus on the most important issue at any given time. This choice involves the ability to integrate all inputs, evaluate them, choose one to concentrate on, and be able to apply our higher-order module to it. This ability is a pretty good correlate of **consciousness**, rather than of simply intelligence.

3.3 An analogy

Humans are left-handed or right-handed. This is because, given a fixed amount of computational power devoted to the both of the hands combined, it makes sense to divide it unequally, so the dominant hand can be used for **all** tasks that require precision.

In the same way, it makes sense, given a fixed brain throughput, for the brain to be able to concentrate on the most important task at a time. This is accomplished by putting resources into a strong central higher-order logic system, which can then be put to various applications as our attention demands.

4 How the neuronal circuitry of the brain leads to consciousness

4.1 A definition of the computational complexity of a neural network

4.1.1 VC dimension

As mentioned, intelligence is the ability to solve problems; the wider range of problems that can be solved, the higher the intelligence. The VC dimension of a neuronal net is a reasonable proxy for intelligence. It tells us what range of problems we are capable of solving. To make it more useful, we should probably impose a metric on the universe of problems – solving a large number of *similar* problems is different from being able to solve the same number of problems that are quite different (but are still amenable to use of the same higher order circuitry).

Perhaps one choice for the metric for the distance between two types of problems is the number of extra layers of a feed-forward neural network that can solve both problems, versus the number of layers for the separate neural networks to solve each individual problem separately.

4.1.2 Kolmogorov complexity

The brain is tricky to compare to software – it's the computer, the operating system, the software, and the data. Perhaps a good comparison for consciousness itself would be software that controls other software.

How do we define, and find, the Kolmogorov complexity of this system? What is the relationship between the Kolmogorov complexity of this, and how close the system is to Turing complete? Remember that the Kolmogorov complexity is defined for code, while being Turing complete is a property of a computional system or language, but also that software that manipulates and changes other software is sort of inbetween code and operating system.

Note that the Kolmogorov complexity refers to the minimum number of statements are required for code to accomplish a task. Perhaps an extension would be to replace number of lines of code with the value of some metric of the neural network. Note that this is not the same thing as number of neurons or number of layers or number of connections, exactly. To put it another way: a language/operating system has a bunch of legal sentences or statements, and the Kolmogorov complexity of a program is the minimum number of these statements in a program that can accomplish a given objective. We need to find the building blocks or axioms of what a neural net can do, and then find the Kolmogorov complexity of a neural net in terms of these fundamental blocks.

We would like to come up with a measure of how close a (recursive) neural network is to being Turing complete.

4.2 Optimizing the computational complexity with a cost function

We want to optimize this (generalized with a metric) VC dimension, but we have to do it subject to a cost regularizer.

This cost, as mentioned, is a combination of the brain complexity and the computational energy expanded. Another possible component of the cost function could be the information (that would be encoded in the DNA) that would be required to specify the topology of the brain/neural network.

It is reasonable to posit that creating neural networks with these constraints will lead to separate modules. As we continue to increase the level of organization of the recursive neural networks with these separate components, we start to reach the level of being Turing complete. This might be considered something of a phase transition! As we reach this level, the computational complexity (VC dimension) increases spectacularly. This is an example of self-organized behavior.

Note that computational complexity, such as the VC dimension, is a **payoff**, while a metric of the neural network, such as the number of connections, is a **cost**. We want to optimize computational complexity with respect to this cost.

Another possibility is to look for a measure of the organization of a neural network as we increase the required VC dimension, for the same number of nodes/connections.

4.3 Organization of neural networks

How do we measure how complex, or organized a neural network is? If everything is attached to each other, that's not really organized. We need to study the strength of the connections to decide if the neural network has self-organized into structures.

One way is to see if there is a fractal dimension we can define. Again, this has to take into account the strength of the connections. A fractal dimension would imply lots of structure on all scales, that is, self-organization. And this has possible links to Turing completeness, recursion and also the chaotic, time-irreversible, dissipative behavior.

4.4 Further

4.4.1 Energy used

How does energy scale with number of connections and strength of each connection?

4.4.2 Neural nets and theory of computation

- I don't think we have a good measure of how close a (recursive) neural network is to Turing complete. Need to look this up or come up with one.

- The multi-layer perceptron (MLP) is a universal function approximator, as proven by the Cybenko theorem. However, the proof is not constructive regarding the number of neurons required or the settings of the weights.

- Work by Hava Siegelmann and Eduardo D. Sontag has provided a proof that a specific recurrent architecture with rational valued weights (as opposed to the commonly used floating point approximations) has the full power of a Universal Turing Machine using a finite number of neurons and standard linear connections. They have further shown that the use of irrational values for weights results in a machine with super-Turing power.(from wiki article on artificial neural networks)

- Juergen Schmidhuber does kolmogorov complexity of neural nets
  - `http://www.idsia.ch/~juergen/`
  - `http://www.idsia.ch/~juergen/loconet/nngen.html`

- http://www.idsia.ch/~juergen/onlinepub.html
- Godel machine talks about "optimality" of solutions
  - http://www.idsia.ch/~juergen/gmfaq.html
  - http://www.idsia.ch/~juergen/gmsummary.html
- Leonid Levin, Levin complexity, Cook-Levin theorem
- Mathis and Mozer

4.4.3 Adapting the Theory of Computation

We don't really have a theory of computation for situations where a software program becomes more complex, only when a computer/computational language does. The brain is programming language, software, hardware, and data rolled up into one.

A program that can manipulate other programs may be a useful approach for this.

4.4.4 Some definitions needed!

What is the exact meaning of recursive? What does this adjective apply to? Is fully recursive the same as Turing complete?

4.4.5 Is DNA complexity a legitimate contribution to the cost?

Is the genetic codebase required to specify brain development a significant drain? It could lead to mutations if it's too long and complicated. This could lead to further self-organization of the brain, similar to zebra stripes and fingerprints. Self-similar (or emergent) structures don't need too much information to form.

4.4.6 VC dimension

The VC dimension is in some way an entropy. The optimization is in some way a maximum entropy method. Entropy is associated with dissipative behavior and time irreversibility. (So?) As per Prigogine, England (MIT, evolution of life) entropy (or free energy) rate maximization leads to complexity, structures, so why not here?

4.4.7 Misc

- The David schwab pankaj Mehta paper of deep learning as renormalization ties in with recursive software as consciousness. Renormalization has at its root the fact that it is the same over different scales, and recursion is also similar.
- The Naftali talk is about the primacy of compression in deep learning.
- Also, his paper is relevant to our comment that consciousness is about generalization... One code that knows how to deal with many situations. That is why evolution came up with consciousness in the first place.
- Phase transitions in neural networks

5 Internal Characteristics of Consciousness

5.1 Qualia

The important thing about qualia: Qualia exist if the person having them believes they are having them. The question of "what are qualia?" should be better phrased as, "what does it mean for a person to believe they are experiencing qualia?", and then: "what are the requirements for someone to be 'sophisticated' enough to believe they are having qualia?".

Which means, qualia are about having a sense of self (the "I" that is having the qualia) and a sense of introspection (in the common and the software coding sense).

5.2 Free will, a special quale

Having a discrete higher-order computational engine and time-switching for what it works on is not enough for consciousness (although it may provide qualia). We need to have some sort of free will for consciousness as we know it to be fully realized. This is given by the time-irreversible and chaotic nature the neural network computation as we reach Turing completeness. As it's fully recursive, the ability to predict the result is lost (we can go around in several different loops any number of times).

5.2.1 The arrow of time

We do not know what consciousness would be like if we did not have an arrow of time for memory (that is, having an equal memory of the past and the future), but it would probably be considered sufficiently different to be a distinct entity. In particular, it is difficult to imagine consciousness without free will, and free will without an arrow of time.

The conscious arrow of time certainly arises from entropy of dissipative systems. But do we need the brain to be a dissipative system, the universe to be dissipative, or both? (Sorry, that's an actual question, there's no answer below).

1. Difference between neural nets and software Because neural nets (and brains) are massively connected, they are not as easily time-reversible as software could be (there are fairly simple debuggers that can walk you backwards!). This is related to entropy, of course, and is strongly important in consciousness.

6 Other humans

This aspect of consciousness is more difficult to quantify, but evolution demands that we be able to model and understand other humans, who grew increasingly complex through evolution. We also need to interact and cooperate with them. And they are modeling us – it's possible that this recursion leads to a phase transition to more complex brains. And that to solve this recursive problem requires recursive brain power.

This gives further value to the introspection that is characteristic of consciousness: understanding ourselves helps us understand others (assuming, fairly reasonably, that others are similar to ourselves). For example, anger can be exhibited by creatures with a wide range of degree of consciousness, but the knowledge of what leads to anger (in ourselves or our peers) can be very useful in deciding our own behavior.

Note that the animals we think of as possessing a higher degree of consciousness tend to be social.

6.1 Evolution, other humans, and intelligence

The problem of evolution of intelligence in social animals is a tricky one.

- Other humans are among our biggest evolutionary competitors: for mating, food, or warfare.

- The smarter we get, the smarter they get, the smarter we have to get.
- However, unlike an attribute like the cheetah's speed, intelligence does **not** tail off in terms of fitness vs. cost. As we are social animals and cooperative, the smarter we get, the more advanced our cooperation can get, the more incrementally useful intelligence gets. That is, the marginal advantage of intelligence keeps *increasing* the more intelligent we get.

7 Compression

7.1 Statement of the problem

Instead of looking at neural networks, we consider a more general algorithm that maps us from problems to solutions. The problems come from a wide range of different types – we need a metric to define what "different" means. It might be useful to have a metric to define how close our solution is to optimal, too.

This problem can be considered as one of compression (see Matt Mahoney).

7.2 Relevance to consciousness

What happens to an optimal (in the sense of Kolmogorov complexity?) compression algorithm as the range of problems that need to be solved goes to infinity? The guess is that, in some sense, the program becomes Turing complete. And yes, this doesn't mean anything right now, because Turing completeness refers to languages, while the program becomes... recursive?

Perhaps this optimization leads to some kind of hierarchical framework for the algorithm; with the highest layer relevant to consciousness.

We're looking at some kind of tradeoff between Shannon throughput and the Kolmogorov complexity (and memory storage) of the algorithm, for some kind of fixed accuracy in performance.

To optimize (compress) a simple text, requires a simple algorithm. The more varied in nature the input, the more flexible the algorithm has to be – it has to look at the input, characterize it, adapt to it and decide the best solution. There should be a link between this complexity, and whether or not a logic system/computer is fully recursive. Is there a continuous measure of how recursive a system is? Also, when the input is other people's behaviour, the complexity goes up a lot, and induces a self-reflective nature (

other people's behaviour is sort of a reflection of us, because they are similar systems) which is a further reason to bring in recursion.

We can then try to mathematically characterize the information content of the input (that is, the sensory information of the world around us), the Kolmogorov complexity (or other complexity) of the conscious brain, and the information content of the set of our reactions to the input. And how does this change as the Kolmogorov complexity/closeness to being Turing complete changes? And how does this change as the input has a sort of deeper pattern common to it (corresponding to our higher-level structure being able to solve many different types of problems)?

7.3 Further

- Is the VC dimension relevant to compression algorithms?

8 Misc

- Higher order thought (HOT): Rosenthal
- Michael Graziano and metacognitive theories !!! very close to my ideas.
- Hutter, compression
- Dr. Giulio Tononi, Integrated Information Theory (involves Shannon Entropy)
- Vijay Balasubramanian, neural coding of information, UPenn `http://www.physics.upenn.edu/~vbalasub/Neuroscience.html`
- Mark Changizi 2AI Labs, Boise Idaho
- Vitanyi, Li page 2: optimal descriptive function, what is that?
- theory of lossy compression – read up
- functionalism, emergent dualism, cognitivism, higher order theory