

Please cite as: Browning, H. & Veit, W. (2021). Evolutionary Biology Meets Consciousness: Essay Review of Simona Ginsburg and Eva Jablonka's *The Evolution of the Sensitive Soul*. Preprint.

Check <https://walterveit.com/publications/> for citation details once published

Evolutionary Biology Meets Consciousness

Essay Review of Simona Ginsburg and Eva Jablonka's *The Evolution of the Sensitive Soul*

Heather Browning & Walter Veit

In this essay, we discuss Simona Ginsburg and Eva Jablonka's *The Evolution of the Sensitive Soul* from an interdisciplinary perspective. Constituting perhaps the longest treatise on the evolution of consciousness, Ginsburg and Jablonka unite their expertise in neuroscience and biology to develop a beautifully Darwinian account of the dawning of subjective experience. Though it would be impossible to cover all its content in a short book review, here we provide a critical evaluation of their two key ideas - the role of *Unlimited Associative Learning* in the evolution of, and detection of, consciousness and a metaphysical claim about consciousness as a *mode of being* - in a manner that will hopefully overcome some of the initial resistance of potential readers to tackle a book of this length.

1 Introduction

What is consciousness? Where does it come from? Why does it exist? How does it work? How can we detect it? These are questions with an incredibly rich and deep tradition throughout the philosophy of mind and philosophy of science, and which form the subject matter of Simona Ginsburg and Eva Jablonka's *The Evolution of the Sensitive Soul*. Animal consciousness is an important and relevant emerging research program – one that is of both scientific and ethical interest (Browning 2020) – and it is into this ripe territory that Ginsburg and Jablonka have entered. Although it is common for the word ‘consciousness’ to be used in a thicker sense in human cases, evoking features such as self-awareness and reflection, this book is primarily about

the very emergence of the first sparks of subjectivity - of minimal consciousness. An organism that is conscious is one that has felt experiences: as well as perception, these can include a range of states, such as pain, boredom, hunger, comfort and curiosity. What they all have in common is their experiential component. Here, Ginsburg & Jablonka (hereafter G&J) poetically draw upon Aristotle's notion of the "sensitive soul" to refer to this subjective experience.

In one of the longest treatises on the evolution of consciousness, G&J tackle the problems of when, why, and how biological organisms developed a sense of felt subjective experience - a point of view - or what they call a new mode of being. In general, G&J see the explanatory project for consciousness as containing three closely linked questions - what it is, how it works, and why it evolved - which must be answered together, with progress on any one question informing and expanding the others in a sort of ever-evolving 'bootstrapping' that allows the eventual explanation to progress. Following a detailed historical overview of scientific and philosophical work on consciousness, G&J introduce the two core ideas of their book: one regarding the metaphysics of consciousness; the other its measurement. It would be an extraordinary challenge to do justice to a book of such breath, nuance, and depth and we will therefore restrict ourselves to a critical discussion of these two novel claims – the empirical claim about the role of *Unlimited Associative Learning* in the evolution of, and detection of, consciousness and the metaphysical claim about consciousness as a *mode of being*.

2 Unlimited Associative Learning

The primary distinctive account contained within this book regards Unlimited Associative Learning (UAL). UAL is "open-ended learning that enables an organism to ascribe motivational value to a compound stimulus or action and use it as the basis for future learning" (p. 191). It is contrasted to limited associative learning (LAL), that can only respond to simple (noncompound) stimuli or actions. UAL involves the perception of multimodal stimuli, fusing different sensory inputs (e.g. colour, sound) into a single percept, as well as the performance of complex actions, the assignation of positive or negative valence to the experience, and the ability to remember and generalise these to other relevantly similar situations. Compared with simpler forms of learning, this allows for greater ability to discriminate stimuli and predict events and most importantly will vastly increase the number of potential associations between stimuli and actions. Performance of UAL requires hierarchical processing, integration, memory and a valenced reinforcement system.

G&J argue that learning plays an important evolutionary role, as it allows ontogenetic adaptation to environments that can occur over a much shorter timescale than genetic or evolutionary change. They propose that UAL is likely to have arisen from the unique conditions of the Cambrian explosion – CNS complexity, environmental change, changes in body plans and increase in complex interactions between organisms (particularly predation), all leading to an evolutionary ‘arms race’ in which improvements in learning capacity gave a selective advantage through behavioural plasticity.

G&J make two distinct claims regarding UAL. The first is that UAL serves as a biomarker for the transition to minimal consciousness. A transition marker is some capacity or component that represents a system having passed a threshold beyond which all can agree the transition has occurred and they argue that UAL plays this role for consciousness. Having a measurable marker such as this allows identification of the presence and distribution of consciousness. This can help us in our understanding of the features of consciousness and the mechanisms by which it operates. Their claim is that UAL serves as a sufficient indicator for the possession of consciousness, i.e. that those organisms that possess UAL are conscious. They argue that UAL is an appropriate biomarker, as it meets the conditions for consciousness. In Chapter 3, they list what they take to be the current scientific consensus on the set of necessary and sufficient attributes of consciousness:

1. Global activity and accessibility (making information available to a number of different cognitive processes)
2. Binding and unification (creation of a single integrated experience from a variety of perceptual and cognitive processes)
3. Selection, plasticity, learning and attention (ability to select between different neurons and pathways/processes and inhibit attention to unnecessary information)
4. Intentionality (aboutness – reference to states of the body or world)
5. Temporal ‘thickness’ (persistence of experiences through time)
6. Values, emotions and goals (the felt valence of experience and subsequent motivation)
7. Embodiment, agency and a notion of ‘self’ (a distinction between self and environment)

They then go on to argue that UAL requires all of these capacities, and thus any organism capable of UAL must possess the appropriate enabling system for consciousness. We can then reverse-engineer from the presence of UAL to the presence of an enabling system that suffices for consciousness.

The benefit to this view is that it has strong theoretical and indirect empirical support, as detailed in the book. It provides a tangible, measurable marker that could be used to demonstrate the presence of consciousness. However, there are several limitations. The first is that UAL is only a *positive marker* for consciousness; the presence of UAL simply represents possession of what G&J consider the necessary components of consciousness. While (if they are correct) any organism capable of UAL must possess the enabling system for consciousness, it is unclear what to say in cases when UAL is absent. We cannot infer that there is no consciousness, though G&J consider it unlikely to be present. Here, they think it depends on the case - in early developmental stages of organisms that will later develop UAL, they are likely to already possess the necessary enabling systems for consciousness, while organisms that never develop this ability are far less likely to. This represents a difference between ontogenetic and phylogenetic development – that is, the development of an individual organism throughout its lifetime, and of a lineage across evolutionary time. Features absent early in the phylogenetic tree need not be absent in the early developmental stages of ‘higher’ creatures, even where these organisms are still just as simple, as they may still possess the necessary functional architecture for their future development. Thus even if juveniles are incapable of UAL, they may still be conscious. In organisms that will never develop this ability, they consider it much less probable.

It is also possible that UAL is too demanding - that though all creatures with UAL will be conscious, there could be some simpler process (such as LAL) that is also sufficient to demonstrate the presence of consciousness. G&J resist this, arguing that the less sophisticated forms of learning do not require the same enabling system. LAL can occur in much simpler brains and thus its performance does not indicate the presence of these systems. Additionally, there is already evidence of LAL occurring without consciousness (e.g. in humans). However, although it is insufficient for attribution of minimal consciousness, according to the criteria they have defined, they do not rule that this means no consciousness is present. It is still possible that a creature capable of LAL possesses the enabling systems for consciousness, it is just that the performance of LAL is not evidence of this.

Another limitation is that, because of its basis in evolution, UAL is only a useful indicator for assessing consciousness in biological systems. G&J are clear throughout that their account is simply one of consciousness as it has actually emerged in biological life on Earth, and they wish to make no strong claims about the possibility of or features of different types of consciousness, such as may be created in machines. As biological organisms are the only beings that we know can be conscious, they wish to remain agnostic about the possibilities of other types of or realisers of

consciousness. G&J allow that artificial systems could be built that are capable of UAL, yet lack the enabling systems for consciousness. However, they do not rule out the possibility of constructing conscious machines, but point out that they would likely require highly complex processing architecture and embodiment. They see consciousness within the biological world as necessarily embodied feelings that are not properties of brains alone, but of embodied brains interacting with an external world via senses and actions.

The primary weakness in their account is that while this may be a convincing theoretical case, with supporting circumstantial evidence, there is currently little empirical evidence *directly* linking the possession of UAL to consciousness (a point also noted by Birch 2020). Importantly, testing is needed to examine whether unconscious UAL is possible. If it is, we can no longer consider UAL to be a marker of consciousness. Of course, here we run into the problem of validation - we need first to determine whether or not an animal is conscious before we can tell whether or not UAL is linked to consciousness, and we have no independent way to do this. Although we can also test for the unconscious performance of UAL, through seeing whether animals (such as humans) we know to be conscious can perform UAL without consciousness (such as through masking experiments), this will not get us far in refuting the account. Even if UAL was performed unconsciously by these subjects, it could be insufficient to reject its link to consciousness. After all, G&J merely argue that UAL is a marker of the presence of the enabling systems for consciousness; even if it is performed without active consciousness, it does not follow that it can be performed in the absence of the necessary architecture for consciousness. In this case, it would still be true that only conscious animals could perform UAL, even if they can do so unconsciously at times.

Here, the precise details of the role of consciousness itself (as opposed to its component attributes) in enabling UAL will become important. G&J link UAL to experienced mental representations – what they call ‘categorising sensory states’ (CSS) – giving rise to learned associations and motivations. These CSSs are generated by the process of UAL and function to both integrate multimodal perceptual experience (of both the environment and one’s own actions) but also to evaluate and reinforce particular action responses through experience of pleasure or displeasure. They consider these to be identical with mental representations, in the typically understood experiential sense of the term. Thus, the representations generated through the process of UAL are necessarily subjectively experienced.

Consciousness, they argue, allows for ‘loading the dice’ in attention and action selection. When an individual is capable of perceiving and remembering an almost unlimited number of compound percepts and actions, this leads to a problem of

appropriate selection - how to effectively choose between competing inputs and motivations (i.e. attention). Loading the dice allows for the most important percepts and actions to take the forefront, through the mechanisms of neural integration, memory and evaluative learning of the relationships between complex stimuli and action responses. G&J take this set of processes to simply constitute consciousness - “the neural dynamics that enable the functioning of complex perception and action ... *is* loading the dice — it *is* minimal consciousness. It *is* what renders an animal sentient” (p. 350). Here, it would not be possible for UAL to occur without consciousness as by definition, the enabling processes for UAL constitute consciousness.

Further empirical work examining whether and how the components of UAL and attributes of consciousness can come apart or work independently, will help provide support for (or refutation of) this view. However, we find that most of the evidence they marshal for the first part of the claim, though indirect, is convincing. The cognitive and neural architecture that is likely to be required for UAL appears to meet the sufficient conditions for consciousness. UAL can be detected both through behavioural ability, as well as the presence of appropriate functional-neural architecture and if these are found to correlate (as G&J contend that they do), there is a strong case for the presence of consciousness. We are more confident than Birch (2020) who considers the proposal “very tentative”: to us, it seems highly likely that creatures capable of UAL will be minimally conscious, though of course further testing will be needed to corroborate this.

More controversial is the second part of this claim, which is more implicit throughout their discussion - that UAL is the driving force for the emergence of consciousness. Different from the ‘backward-looking’ claim that the presence of UAL indicates the presence of a suitable enabling system for consciousness, this is a ‘forward-looking’ claim that consciousness facilitates the performance of UAL, and provides benefits to the organism. Although they argue that consciousness is not a trait but a mode of being, and thus does not have a function, they also appear to concede that it has benefits in facilitating particular types of learning and decision-making. In particular, when discussing CSSs, they think that the action-guiding and motivational force of these mental representations will provide an adaptive advantage to organisms that possess them.

Here, we must be careful to disentangle claims about the emergence and evolution of consciousness from those about UAL. It is indisputable that UAL would provide an adaptive advantage. G&J argue that UAL is beneficial for organisms: it allows for a huge number of potential learned associations, facilitates higher order learning and allows for open-ended adjustments through cumulative learning and

development of complex behaviour. This gives rise to a high degree of plasticity throughout a lifetime, analogous to the way in which heredity allows for change over evolutionary time. They consider associative learning to have been one of the drivers of the rapid diversification during the Cambrian, while the complex changes and conditions at this time also gave rise to an environment conducive to UAL, driving evolution in a positive feedback loop. This force was so strong that they think it was likely to have driven two or three separate origins of UAL and minimal consciousness. This is a convincing story about the *evolution of learning*, but does it tell us anything about the evolution of consciousness itself? In particular, what can it tell us about the evolved link between UAL and consciousness?

One way to conceive of this is that the UAL architecture is identical with that required for consciousness, but this again precludes the possibility of consciousness as occurring separately from UAL, which they seem to want to allow. So we should think that there is both an UAL architecture, and an enabling system for consciousness which, while in some way dependent on one another, are not identical. This then gives rise to two possibilities: The first is that selection for UAL drove the development of the necessary neural systems, which gave rise to consciousness. The second is that the enabling systems for consciousness, already in place already for some other reason, allowed for the evolution of UAL.

It is not always clear which direction G&J wish to take. In describing UAL as a biomarker for consciousness, they appear to take the second view - that while UAL is sufficient for consciousness, it is not necessary for it. It is the presence of the features of consciousness that allows for UAL. However, in outlining their evolutionary story for UAL, it appears much more like they take the presence of UAL architecture as necessary for consciousness. They also expect that a loss of UAL ability in a lineage will also eventually lead to a loss of consciousness, as the enabling system is costly and would no longer be beneficial. This obviously assumes that consciousness could not benefit organisms in other ways such that the enabling systems could remain in place. If these systems arose for some other function, eventually being co-opted for UAL, then it is not clear that they would be lost if UAL was. It is also possible that early consciousness did not even resemble current minimal consciousness with the exception of a minimal sense of a (very different) point of view. When considering the evolution of life, Dennett (2017) argues that the first reproducers, rather than being very simple replicator molecules, may actually have been more complicated and ungainly mechanisms that over time simplified in competition with others. It is possible that consciousness could be the same - that the conscious processes we see today are themselves simplified and streamlined versions of what were initially laborious and complex processes that may have

worked quite differently and served different functions. Although this work provides a convincing case for a strong link between the evolution of UAL and consciousness, its details remain unclear. We need not think that the selective benefits of UAL are themselves explanatory of the evolution of consciousness.

3 Modes of Being

The second novel idea within their book is to conceive of consciousness as a new *mode of being*, rather than a mere trait. This part of their argument may appear unusual to many operating in the debate, not the least because this formulation - not unlike their choice to include Aristotle's sensitive soul in the title - evokes a sense of outdated and strange metaphysics. We share some of this opposition to this vocabulary, but think it best conceived as a metaphor.

They begin their book by introducing the idea of teleological (goal-directed) systems and the three 'modes of being', taken from the works of Aristotle, each of which is considered to have a unique telos (goal). These are: life (survival/reproduction), sentience (value ascription to stimuli), and rationality (value ascription to concepts). The focus of this book is the second of these - the "sensitive soul". Rather than a trait, such as vision, **G&J** see consciousness as a *mode of being*, in the same way as the emergence of life and rational thought also constitute new *modes of being*.

In several places throughout their book, G&J motivate their account through this analogy, i.e. by drawing a parallel from consciousness to life and/or rationality. Neither, they think, can be captured in a simple definition or trait thus explaining the lack of progress on trying to come up with definitions for these phenomena. Compare their discussion of the distinction between life and non-life. Life, they argue, is not a functional trait that organisms possess, but rather a new way of being that opens up new possibilities; so too with consciousness. It is a new form of biological organization at a level above the organism that gives rise to a "new type of goal-directed system", one which faces a unique set of challenges and opportunities. They identify three such transitions - the transition from non-life to life (the "nutritive soul"), the transition from non-conscious to conscious (the "sensitive soul") and the transition from non-rational to rational (the "rational soul"). All three transitions mark a change to a new form of being, one in which the types of goals change. But while this is certainly correct in the sense of constituting a radical transformation in the kind of goal-directed systems there are, we have qualms with the idea that this formal equivalence or abstract similarity can be used to ground more concrete

properties. Yet G&J use this analogy to motivate their UAL account in parallel to *unlimited heredity* as a transition marker of life.

What G&J want to convey is the idea that a lot of things have to be in place and fine-tuned to each other in order to enable (potentially) unlimited heredity. These conditions are not easily achieved, but once they are in place, this new mode of material existence can rapidly change the face of the Earth. So too, they argue, is consciousness something that requires the manifestation of a certain number of other traits, but once in place can cause a rapid change, as seen in the Cambrian explosion. This leads G&J to offer a provisional account of consciousness as “a mode of being that involves activities that generate temporally persistent, dynamic, integrated, and embodied neurophysiological states that ascribe values to complex stimuli emanating from the external world, from the body, and from bodily actions” (p. 7). To see consciousness as a complex bundle of traits may not be best described as a *mode of being*, but it highlights the fact that the living world undergoes a rapid transition once everything falls into place.

Now, while this way of putting things may appear beneficial from a methodological perspective, G&J deny that they see these three forms of goal-directed systems as merely convenient ways of speaking about them. They hold that these transitions constitute a real emergence of telos - systems that are “intrinsically teleological” - rather than just useful ways of thinking. If successful, this approach would gain much support simply in virtue of functioning as an explanation for one of the most puzzling phenomena in evolutionary time. Indeed, the Cambrian is now commonly associated among those interested in the evolution of consciousness or ‘subjectivity’. Like Godfrey-Smith (2016) we may wish to recognize that there is a transition somewhere in the Cambrian from which minds “evolved in response to *other* minds” (p.63), though it might then be more parsimonious to recognize the Cambrian as an explosion or emergence of the capacity to represent the mind (and knowledge states) of others (Veit, forthcoming).

One further concern is the use of only three distinctions in the modes of being (cf Dennett’s (1995) analogous but more fine-grained distinctions between Darwinian, Skinnerian, Popperian and Gregorian creatures). But such a more fine-grained picture doesn’t allow for the neat parallels between the “three souls”. If one doesn’t force the idea of modes of being into a three-fold picture with formal equivalences between each transition, we may be able to recognize many more new types of goal-directedness, arising from the different types of goals and, we suggest, each potentially giving rise to different types of normativity. We are somewhat worried that their attempt to draw on the research program on the origins of life may eventually limit innovative approaches if the similarities are taken too seriously.

Nevertheless, we agree that the nutritive soul has goals simply relating to survival and reproduction, and normativity here arises at the level of biological functioning - things can go better or worse *with* these organisms. The sensitive soul adds a layer of additional goals relating to felt experience - things can now go better or worse *for* the organism, and they will act for subjectively felt reasons at the proximate level, even if they are still grounded in fitness at the ultimate level. This gives rise to *wellbeing* and the normativity associated with pleasure and suffering. Finally, the rational soul provides the ability to deliberate and consider these reasons, to take them as things in themselves. Here, normativity arises in its most complex form, in terms of the considered preferences of agents. Dennett argues that distinctively human consciousness has allowed humans to become “the only species that has managed to occupy a perspective that displaces genetic fitness as the highest purpose” (Dennett 2017).

On a purely methodological level, their approach offers a neat instruction for how to approach the problem of consciousness, i.e. by following in the footsteps of research on the origins of life and trying to learn from their mistakes and successes. Similar to research on the origins of life, research into the origins of consciousness becomes transformed in the search for the dynamics, conditions, and organisational principles necessary for subjective experience - making UAL a transition marker for consciousness in a similar manner to how some astrobiologists have embraced unlimited heredity as a transition marker for life. The very same argumentative strategy could also be used in future to develop a similar account of the rational mode of being, as they sketch out in the end of the book, relying on open-ended symbolic representation.

Nevertheless, we worry that G&J treat these parallels too optimistically. Rather than thinking of these parallels as a highly useful and novel way of thinking about an old problem by drawing on the tools developed for solving another (something that is not unusual in science), G&J often appear to be too committed to their general framework which can lead to overstating their conclusions. This may not be too problematic, so long as we consider it as an ambitious attempt to open the doors for a new research program in the science of consciousness - and this should only be praised. For sciences to progress, we need something like paradigms in a weak sense for scientists to work in. That their account could turn out to be wrong is a strength, not a weakness, though further work is required to create a testable program - something that can tell us which animals are ‘participating’ in these new ways of being and which aren’t.

4 Conclusion

This review is only a rough attempt to fully cover the richness found in *The Evolution of the Sensitive Soul*. This book contains what appear to be many sidenotes and diversions, particularly the exploration and celebration of the lives and works of particular thinkers the authors have deemed notable, such as Yeshayahu Leibowitz, an Israeli intellectual and early philosopher of biology who argued for an explanatory gap between the mechanistic and the teleological and Jean-Baptiste Lamarck, who they consider to have been “scandalously misrepresented” (45) within biology. However, these are all part of their method of building their account on the work of philosophers and scientists that have come before them - the teleological frameworks of Aristotle and Dennett, the reductionist evolutionary psychology of Lamarck, the ‘stream of consciousness’ metaphor of William James, to name just a few. It an approach that Schliesser (2019) dubbed *Synthetic Philosophy*, a style of doing philosophy that synthesizes evidence from different disciplines in order to make sense of an as of yet poorly understood phenomena, and is similarly found in works like Daniel Dennett’s (2017) *From Bacteria to Bach and Back* and Peter Godfrey-Smith’s (2016) *Other Minds*.

It is rare to find a book of this size that has such a meticulously crafted and convincing argumentative road such as this. Even less so when it integrates findings across incredibly diverse fields such as neuroscience, evolutionary biology, psychology, cognitive science, and the respective histories of these fields. They provide a depth of scientific detail in support of their account. Ranging from evolutionary to cellular biology, G&J’s biological expertise is clear throughout. This is not always for the faint-hearted – even with a background in biology, one can struggle to grasp all the details – and G&J are cognizant of this, allowing that readers may skip the technical details they are less interested in. However, overall it is heartening to see so much evidence martialled in defence of what could otherwise be seen as just another set of speculative claims about the evolution of consciousness.

With a book of this size it becomes a difficult task to hold the reader’s attention and avoid drifting away into unnecessary detail; a challenge Ginsburg and Jablonka master with surprising ease. This is achieved in part through their use of magnificent illustrations. It is not often that a science or philosophy book is also artistic, but the playful and surreal illustrations by Anna Zeligowski add an aesthetic dimension to the reading of the book that adds support to those who seek to recognize the important role of aesthetics within science. The ‘dreaming elephant’ on page 448 was a particular favourite of ours, evoking Kipling’s *Just So Stories*. Though they are justifiably worried that their greatest challenge will not be to convince people of their

arguments but rather to read a book of this size to begin with, it is our hope that this review will help others to overcome this initial hurdle and motivate them to join the efforts to explain the evolution of consciousness.

What we have attempted here, is to offer a brief analysis of the two key ideas in a book that is about to leave a tremendous impact on the study of consciousness, yet cannot be neatly characterized as either philosophy or science. It is *both*. Indeed, we may go further and recognize a genuine milestone. Whether or not one ultimately agrees with the account presented within the book, the detail in which it is presented makes this essential reading for anyone working in, or just interested in, the evolution and workings of consciousness.

References

- Birch, J. (2020). In search of the origins of consciousness. *Acta Biotheoretica* (68), 287–294.
- Browning, H. (2020). If I could talk to the animals: Measuring subjective animal welfare. PhD Thesis, Australian National University. <<https://openresearch-repository.anu.edu.au/bitstream/1885/206204/1/Browning%20Thesis%202020.pdf>>
- Dennett, D. C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. Simon and Schuster.
- Dennett, D. C. (2017). *From Bacteria to Bach and Back: The Evolution of Minds*. WW Norton & Company.
- Godfrey-Smith, P. (2016). *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness*. Farrar, Straus and Giroux.
- Schliesser, E. (2019). Synthetic philosophy. *Biology & Philosophy* 34(2), 1–9.
- Veit, W. (forthcoming). The evolution of knowledge during the Cambrian explosion. *Behavioral and Brain Sciences*. <<https://doi.org/10.31234/osf.io/2rm8k>>