Instability physics: Consciousness and collapse of the wave function

Alex Hankey
Swami Vivekananda Yoga Anusandhana Samsthana
Eknath Bhavan, 19 Gavipuram Circle
Kempegowda Nagar, Bengaluru 560019, India
alexhankey@gmailcom

Abstract Roger Penrose has long supported the view that wave function collapse, or more accurately wave packet reduction, is one of the defining aspects of consciousness. In collaboration with Stuart Hameroff, he has suggested that it can be achieved by quantum gravity, giving rise to their well-known Orch OR theory of wave function collapse. Their theory has a fundamental disadvantage, however; Quantum gravity is too general; it applies at all points where matter is present, i.e. in all kinds of circumstance where wave packet reduction is not observed. There is no hint as to why cell interiors are of special importance. That assumption cannot be justified. However, the Orch OR approach does acknowledge that only a situation governed by purely non-linear physics can offer circumstances where quantum wave-packets will be reduced; non-linearity must be at the heart of the theory. This paper identifies a biological condition of non-linearity providing an advantageous way to explain objective reduction, OR, of wave packets: critical instability in complexity biology, known as ‘criticality’. The condition provides a rigorous reason why wave packet reduction should occur; one grounded in well-accepted modern biology, complexity biology. This paper develops the idea further: it demonstrates how critical instabilities are found not in the cytoplasmic reticulum where Penrose and Hameroff site their Orch OR, but under the very circumstances where one would expect consciousness to be located: at the loci of control of biological regulatory systems. Wave packet reduction can then select the future that the organism desires, doing so on a microscopic quantum level.

1 Introduction
The main thread in 20th century attempts to elucidate how the human brain supports consciousness originated with one of the major advances of 20th century mathematics, the development of metamathematics in 1931[1] by Austrian mathematician, Kurt Gödel. Gödel proved that no system of mathematics that includes arithmetic can be considered complete, in the sense that further statements that are true can always be found. Such statements are perfectly valid, but cannot be proved from within the system of axioms already in use.

Gödel’s method of establishing the existence of statements that are true but unprovable requires ‘stepping outside’ the system of axioms. The Oxford philosopher, JR Lucas [2], suggested that the human mind must therefore operate outside any system of axioms. Although it can use them when it so chooses, its inherent function does not involve mathematical algorithms. After making this sweet suggestion, Lucas found himself opposed by a barrage of protests from materialist academics [3], who were happy to work on the limited assumption that human minds are merely material machines that have no genuine
‘imagination’, in the sense of conjuring up ideas not based on logical deduction Rather, mind works on the basis of fixed algorithms

Lucas’s insight, and the opposition arising from it, stimulated the Oxford mathematician and physicist, Roger Penrose to start work establishing a theory of mind and consciousness His starting point was metamathematics; the result, the Emperor’s New Mind [4] The academic community’s reaction to his efforts was as negative as to Lucas’s: Interesting ideas, but no proof! No one was persuaded to give up their materialist prejudices

One person who showed deep interest in Penrose’s goal of establishing a theory of conscious awareness was the anaesthesiologist, Stuart Hameroff, whose book, Ultimate Computing [5], explored possible roles for cell microtubules in supporting consciousness Microtubules have pervasive roles in cellular operations and may support computational functions Might they be sufficiently complex to somehow support consciousness?

Penrose decided to use his skills as a physicist He had worked extensively with his Cambridge colleague, Stephen Hawking, to prove a theorem about the end and origin of space-time [6], a contracting universe will end in a point, a mathematical ‘singularity’ He turned to physics in his attempt to establish a rigorous theory of consciousness

As is well known, quantum physics [7] is a field where a non-trivial role for consciousness is hypothesized [8] It is totally different from classical physics which represents the physical bodies that we perceive as continuously existing objects – just as our brains interpret them Its fundamental concepts are quanta, like electrons and protons, and ‘wave functions’ that it uses to represent them Quanta exist in free motion and in bound states such as electrons in atoms When quanta in free motion interact with matter, they may get captured Electrons enter into bound states, while photons give up their energy causing excited states of various kinds When such an event occurs, the wave function of the quantum is said to ‘collapse’ [8,9] The process is more accurately described in terms of ‘mixtures of wave functions’ [8] known as wave-packets, as ‘wave-packet reduction’ (WPR) [8] Measurements of quanta usually require a quantum to interact with a physical apparatus, so they involve WPR

A major problem with quantum physics is that, while it describes free motion and bound states, it cannot describe the WPR process It describes some of the physics of quantum particles like electrons, or photons, particles of light, but not the measurements made on them It can describe many ways that quanta behave, but not the WPR process involved in measuring them [8,9]

The WPR process lies outside the scope of quantum physics Nor is it to be explained by classical physics describing our world of sense perception Yet it produces the information that experimenters obtain concerning quantum systems It results in information in the minds of those operating the apparatus in each experiment

How can this be? It was surmised that the cause of wave function collapse [9], or wave packet reduction, is the consciousness of scientists conducting the experiment Observers see and record definite states, not continuing mixtures [8] Could the observer’s consciousness somehow be responsible for WPR? Should that be the case, the physics underlying conscious experience in the human mind should also be responsible for reducing wave packets

Physicists are generally willing to agree (or at least acquiesce) that a defining aspect of any proposed physics of consciousness must be its ability to reduce wave packets [8] However, classical physics cannot do it; quantum physics is equally impotent: so, the physics responsible for WPR must lie outside the realm of both quantum and classical physics

This important statement / theorem has largely been lost on many scientists who have tried to solve the problem It has proved highly intractable After all, what other physics can there be?

It was not recognised by Penrose After ‘The Emperor’s New Mind’ [4] failed to convince the world of science that consciousness is not an algorithm-using mechanism, he began collaborating with anaesthesiologist, Stuart Hameroff, whose book, Ultimate Computing [14], explored the possibility that cell microtubules known to have pervasive roles in cellular operations, might be sufficiently complex to somehow support consciousness The idea was that WPR occurs in the microtubules, so they should be involved in consciousness
Penrose then wrote a second book, ‘Shadows of the Mind’[10], suggesting that consciousness actually has its basis in WPR. He suggested that it is instantiated by that process, existing in a succession of instants of cognition. His paper with Hameroff [11], suggesting a form of WPR in brain microtubules, followed.

Again, strong objections were made [12] to Penrose’s new program; The physics community remained unimpressed. Penrose was persuaded to participate in a debate at Cambridge with his former colleague, Stephen Hawking, and US scientists Abner Shimony and Nancy Cartwright. Once again, Penrose’ ideas were rejected [13].

So, the problem of how consciousness can cause WPR remained unsolved. Another intractable problem is the question of how consciousness has self-knowledge. Thirty years ago, ‘That which has Self-knowledge’, was the standard definition of the word, ‘consciousness’, given in English dictionaries [14]. We may legitimately inquire what physics can make abstract self-knowledge possible?

This paper shows that the two problems are deeply interrelated and have a common solution. Starting from a concept originating in complexity biology, criticality in regulation, it shows how that implies the existence of perfectly self-observing information loops [15]. This concept is then used to yield a common solution to both problems.

2 Methods

The approach used here is to exploit the presence of critical instabilities in biological control systems first shown in studies of complexity biology [16,17]. Complexity biology’s sub-discipline of fractal physiology [18] found that patterns of response of all physiological systems to sequences of fixed stimuli are anomalous. Instead of fixed responses to sequences of fixed stimuli (give or take variations caused by thermal fluctuations), physiological systems yield entirely different patterns of response. Standard distributions of responses given by physiological (and psychological) systems are fractal (1/f) distributions [19], known to be characteristic of critical instabilities in other scientific fields, particularly those connected to phase transitions [20].

As shown in the sections that follow, the physics of instability can be used to explain both WPR and self-knowledge.

3 Results

(1) The first important result derives from showing that critical instabilities contain an infinitesimal ‘perfectly self-observing loop’ of information. This occurs in the case of critical instability in flowing fluids, which occur at the critical Reynolds number [21,22].

In fluid flow, the information in laminar flow consists of well-defined flow vectors at each point in the fluid. Above the critical point, vortices occur, causing flow vectors to change as if randomly at each point; but at the critical instability, the vortices are only infinitesimal, and cannot manifest fully. At a fluid’s critical instability, its flow vectors may be held to gain infinitesimal vortices attached to them. This constitutes the first step [23,24].

(2) Because the information in flowing fluids presents as flow vectors, the infinitesimal vortices attached to the flow vectors at critical Reynolds number can effectively be considered, ‘information loops’. This constitutes the second step [24].

(3) All information loops can be assigned a gain, g [25]. For infinitesimal loops, the gain can be identified as 1, g = 1. Such a loop may be considered ‘perfectly self-observing’ [23].

(4) The presence of infinitesimal information loops may be hypothesized to generalise to other systems, i.e., to other forms of critical instability [24].

(5) Because observation causes WPR, the presence of an infinitesimal, perfectly self-observing information loop collapses all wave functions in the system, i.e., it causes internal WPR [23]. Result 5 constitutes the First Main Result Critical Instability causes internal WPR in a system.

(6) Such WPR is also the root cause of the instability [24]. Systems are stabilized by the presence of quanta, like phonons that keep the density of solids and liquids stable. Should its stabilizing quanta be eliminated, a system will become unstable. WPR reduces all wave packets, eliminating stabilising quanta.
(7) It thus appears that critical instability and WPR are two sides of the same coin. They go hand in hand: one cannot be present without the other. Use WPR to remove internal stabilizing quantum fields: instability results. Equally, instability can be attributed to loops of information that are infinitesimal and therefore perfectly self-observing – WPR results.

Now extend the process of WPR to wave functions or wave packets impinging on the system and causing excitations therein. Their quanta will be unable to generate excitations of a quantum nature within the system, because the instability based WPR will destroy such excitations. Thus, when the system is raised to excited states by interacting with quantum wave functions or wave packets in the external world, the system will collapse or reduce any quanta that attempt to manifest in the system WPR extends to external quanta.

Now to the final step, (and the Second Main Result):

(8) A perfectly self-observing information loop confers consciousness on a system [23] This can be understood by considering the subjective experience of being in a state of pure consciousness with no thoughts present. That occurs when the mind is cleared of all excitations, i.e., thoughts and feelings. The result is a state where ‘the Self knows Itself alone’. Known as ‘pure consciousness’, it is well explained in the ancient Indic literature, and well understood throughout the sub-continent.

But in physical terms, how can there be a state where ‘the Self knows Itself alone’? Moreover, a state that is also the ground state of states of awareness with content?

From our analysis above, critical instability seems to offer a basic solution. An infinitesimal, perfectly self-observing information loop functions like ‘a self-knowing itself’. Clearly, the information flow in the loop is ‘informing itself’. What does it tell it? Nothing, no measurable property (no thoughts), no qualia (no feelings). But: it provides a basis of self-reference for any information added on top of it.

Any sensory or mental information added on top of the state of ground will be treated as content of personal experience, i.e., as ‘my information’ to the being identifying with that loop.

This leads to the Second Main Result: An infinitesimal self-observing loop can model the Self in the state of pure consciousness, i.e., where ‘the Self knows Itself alone’.

Combining the Two Main Results yields the Thesis proposed in the Title of the Paper. Add to the first main result (the WPR ability of instabilities), the second main result (the ability of perfectly self-observing loops to generate a sense of Self-knowing Itself, and thus form the basis for conscious experience) We then obtain the Third Main Result:

(9) Perfectly self-observing loops generated at critical instability can both reduce wave packets and generate a sense of Self-knowing Itself with capacities for mental and sensory experience i.e., they explain how ‘Consciousness is responsible for wave function collapse / WPR’ QED.

4 Discussion

In biological organisms, the capacity for experience is based on the perfectly self-observing information loops produced by critical instabilities at the apex of biological regulatory systems. Such critical instabilities occur precisely where one would expect consciousness to be located: at the regulatory systems Loci of Control, rather than in the cytoplasmic reticulum where Hameroff and Penrose sited Orchestrated Objective Reduction (Orch OR) [11], to relate to possible mechanisms of anaesthesia.

A potential for future discussion is: From loci of control, it is natural for WPR to make choices selecting the organism’s desired future [21]. The fractal distributions of response make this an attractive topic for detailed investigation. Also, as proposed by Nobelist J C Eccles, dualism is severely wounded, but not diseased [26-28]. The imminent discovery of higher dimensions may reopen this conundrum.

The new theory has the capacity to show how experience, consciousness, and the capacity to choose are all bound up in a single physical circumstance, perfectly self-observing information loops at critical instability. It supports the idea that consciousness is generated at a microscopic level, in cells, as Hameroff wishes to do, and in nervous systems, for which states of critical instability are also fundamental in fractal physiology [18].

Penrose was aware of the need for non-linear physics to reduce wave packets. Being expert in...
quantum gravity as shown in his book, The Road to Reality [29], and the field in which he had collaborated with Hawking [6], he knew that it incorporates highly non-linear physics and was a candidate for the source of WPR [30].

Recent versions of the Orch OR theory [31] incorporate that idea. But that approach has a fundamental disadvantage. Gravity is too general a phenomenon; ultimately all gravity is quantum gravity. It applies wherever matter is present. WPR would occur not just in cells and brains, but everywhere all the time, in circumstances where it is known not to. The Orch OR theory of WPR does not explain why cell interiors are special. But it implicitly recognizes that only extreme non-linear physics can cause WPR. In this, it is in the right direction. But the theory described here shows that the extreme non-linearity of critical instabilities is required to produce WPR, and, for further reasons to generate abilities for self-awareness, self-knowledge, and consciousness.

Criticality physics from complexity biology has several advantages: (1) it explicitly bridges the gap between quantum and classical realms (as might have been hoped); (2) it contains perfectly self-observing loops, which obviously cause WPR; (3) it uses those loops to generate the Self-knowledge characteristic of consciousness; (4) it is sited at regulatory systems’ Loci of Control, where executive ability must be located; and (5) it offers subjective awareness choice, so that, rather than merely being an impotent observer, it empowers consciousness to make non-trivial decisions.

Strengths: the major strength of the theory lies in it being grounded in well-accepted modern biology, complexity biology. Moreover, its physics integrates feedback from control theory, thermodynamics, and optimization of performance – the reason for criticality’s first adoption during life on earth [17].

A further strength is that at instability, quanta cannot exist [23], and because quanta are responsible for stabilizing material objects on a gross level [23], no gross object can be created. The theory is consistent with consciousness existing at a ‘subtle’ level, lying hidden from the gross senses [32].

5 Conclusion
The physics of critical instability generates Perfectly Self-observing Information Loops that underlie awareness of Self-knowing-Itself and are also responsible for Wave Packet Reduction. Being at loci of control of organism regulation from where motor functions operate, they can be related to motor functions of consciousness. The approach thus explains how ‘consciousness reduces the wave-packet’, or ‘collapses the wave-function’.

Acknowledgements
Development of this theory has benefitted from discussions with many people over many years, not least Stuart Hameroff; also, Brian Josephson, Richard Amoroso, and colleagues in India, notably Dr Vijay Bhatkar, Chancellor of Nalanda University, Dr Judu Ilavarasu at SVYASA, and Dr K Sankaran, Director of Justice KS Hegde Institute of Management, Nitte, Karnataka.

References
[1] Gödel K 1931 Über formal unentscheidbare sätze der principia mathematica und verwandter systeme, Monatshfte fur Mathematik und Physik 38 173–198; Trans Gödel K 1995 Some basic theorems on the foundations of mathematics and their implications; S Feferman ed. 1951 Kurt Gödel Collected works 1995 Vol III 304–323 Oxford Univ Press
[2] Lucas JR 1959 Minds, machines and Gödel Philosophy 36 112-127
[3] Lucas J R 1996 The Gödelian argument, https://wwwopenstartsunitsit/handle/10077/5477
[4] Penrose R 1999 The Emperor’s New Mind: Concerning Computers, Minds and the Laws of Physics Oxford University Press
[5] Hameroff S 1987 Ultimate Computing Elsevier ISBN 978-0444702838
[6] Hawking SW and Penrose R 1970 The singularities of gravitational collapse and cosmology Proc R Soc Lond A 27 3141519 529-48
[7] Dirac P 1958 The Principles of Quantum Mechanics 4th Ed, Oxford University Press, London
[8] Stapp H P 1999 *Mind, Matter and Quantum Mechanics*, New York Springer, 2nd Ed 2013
[9] Heisenberg W 1927 Über den anschaulichen Inhalt der quantentheoretischen kinematik und mechanik, *Z Phys* 43 172–198
[10] Penrose R 1994 *Shadows of the Mind: A Search for the Missing Science of Consciousness* Oxford University Press
[11] Hameroff S Penrose R1996 Orchestrated objective reduction of quantum coherence in brain microtubules: The Orch OR model for consciousness *Math computer simulation* 40 453–480
[12] Grush R Churchland P 1995 Gaps in Penrose's toilings *J Consciousness Studies* 2 1 10–29
[13] Penrose R Shimony A Cartwright N Hawking S 2000 *The Large, the Small and the Human Mind* Cambridge University Press
[14] Oxford English Dictionary 1973
[15] Hankey A 2012 Abstract, *Science of Consciousness Conference* Tucson
[16] Waldrop MM 1994 *Complexity: The Emerging Science at the Edge of Order and Chaos* New York, NY Penguin
[17] Kauffman S 1995 *At Home in the Universe The Search for the Laws of Self-Organization and Complexity* Oxford University Press, New York
[18] Bassingthwaighte J B Liebovitch LS West B J 1994 *Fractal Physiology* Oxford University Press
[19] Bak P Tang C and Wiesenfeld K 1987 Self-organized criticality: An explanation of the 1/f noise *Phys Rev Lett* 59 381–384
[20] Stanley H E 1971 *Introduction to Phase Transitions and Critical Phenomena* Ox Univ Press
[21] Stokes G 1851 On the experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels *Transactions of the Cambridge Philosophical Society* 9 8–106
[22] Reynolds O 1883 An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels *Phil Trans Roy Soc* 174 0 935–982
[23] Hankey A 2014 Complexity biology-based information structures can explain subjectivity, objective reduction of wave packets, and non-computability cosmos and history *J Natural and Social Phil* 10 1 237–250
[24] Hankey A 2015 A complexity basis for phenomenology: How information states at criticality offer a new approach to understanding experience of self, being and time *J Prog Biophys Mol Biol* 10 237–250
[25] Wiener N 1948 *Cybernetics or Control and Communication in Animal and Machine* MIT Press
[26] Amoroso R L 2009 Ed. *The Complementarity of Mind and Body* Nova Science Press
[27] Amoroso R L 2015 Toward a pragmatic science of mind *Quantum Biosys* 6 1 99-114
[28] Amoroso R L and di Biase F 2013 Crossing the psycho-physical bridge: Elucidating the objective character of experience *J Cons Explor & Research* 4 9 932-954
[29] Penrose R 2005 *The Road to Reality: A complete guide to the laws of the universe* Knopf NY
[30] Ellis J Mohanty S and Nanopoulos DV 1989 Quantum Gravity and the collapse of the wave function *Phys Lett B* 85 221 113
[31] Hameroff S Penrose R 2014 Consciousness in the universe: a review of the ‘Orch OR’ theory *Physics of Life Reviews* 11 39–78 & Penrose, Orch OR References therein
[32] Hankey A 2019 A means for liquid water to retain information (this volume)