Consciousness without the Infinite Regress

Consciousness without the Infinite Regress

By

Guenter Albrecht-Buehler

Cover illustration: Inner Matrix (Pentomino mosaic 4/85 by the author)

To my wife Veena, and to our children, Ananya, and Ishan,
And with no less love
To my older children, Christine, Conrad, and Carl.

Matter is not so material and mind not so mental, as is generally supposed.
—Bertrand Russell (1872-1970) in 'The Analysis of Matter'

CONTENTS

SUMMARY

Consciousness had the most profound effects on our biological success – but also on the threat of our species. On the one hand, in the short span of a few million years it accelerated our specialization and communication skills far beyond the levels that genome variation and natural selection alone could have achieved. Thus, it rapidly turned us into the most effectively cooperating species on the planet. On the other hand, it has designed and contrived 'justifications' for acts eviler than were perpetrated by any other species on their fellow members and on others.

So far, consciousness has stubbornly defeated our attempts to explain it as a **product of non-mental elements.** The ultimate reason for this failure is that we **are conscious of being conscious**. As a result, all non-mental-based explanations must lead to endless chains of **infinite regress**.

Therefore, the present book, inspired by Bertrand Russell's reminder that mind may be more material, and matter more mental, than we tend to assume, proposes to drop the requirement of non-mental elements of explanations. Instead, it bases the explanation of consciousness on elements that are already **'borderline-conscious'** like, e.g., **'qualia'**.

This approach tries to move beyond the assumption that consciousness can be explained with model neurons and the electrical and chemical signals that they exchange. Instead, it proposes to submit those 'elementary' signals and their carriers to a first process of **mentation**, which adds 'rules', 'decisions', and 'meanings' before employing them as elements for the explanation of consciousness.

More specifically, and based on additional ideas derived from Schopenhauer,[1] Glaucoma, and hashing algorithms, this book proposes that:

(a) Our mind, seamlessly and with high speed, stitches together an internal image of the world from countless patches of visual,

---

[1] Arthur Schopenhauer (1788-1860) 'Die Welt als Wille und Vorstellung (The World as Will and Representation [1819]).' The literal translation of 'Vorstellung' should be 'Imagination' not 'Representation'. Schopenhauer was highly regarded by many physicists, most notably Albert Einstein, Erwin Schrödinger, and Wolfgang Pauli. Einstein described Schopenhauer's thoughts as a "continual consolation" and called him a genius.

auditory, tactile, olfactory, and spatial information. We will call it the **'inner matrix'**. We continuously scan and interpret the contents of the inner matrix as a necessary condition of our survival, thus generating our so-called **'qualia'**.

(b) Qualia play for our mind the same role that **'hashes'** (short, standardized, unambiguous, and unique tokens of immensely large and complex 'originals') play for our modern information technology.

(c) Based on these assumptions, **consciousness is interpreted as a compound quale**, which integrates continuously the partial qualia that are present in our mind. Analogous to what every modern video game can do, the combined personalized information of this **quale of all qualia** computes a remote self-image that feels like an 'out-of-body' experience of us, and of the world at large.

Ultimately, this book is an attempt to demystify consciousness without trivializing it. Thus, it follows in the footsteps of Daniel Dennett, Susan Greenfield, and Antonio Damasio, who have worked for decades to clear consciousness of semi-religious interpretations. Nevertheless, the students of their work will notice that the present book differs substantially from Dennett's views of 'qualia,' and what is defined as the 'inner matrix'. It is also not centred on electro-physiological experiments like Greenfield's work. Likewise, the image of our inner matrix as described by Damasio is much more detailed and anatomy-oriented than mine. Yet, I hope that these differences will not disqualify me as an eager comrade-in-arms in their battle to keep the discussion about consciousness on sober minded, logical grounds.

PREFACE

HOW CAN DIRTY WATER AND DIRTY TOWELS GET DIRTY DISHES CLEAN?

A Universe of Biological Codes

Consciousness becomes all the more bizarre, and even more threatening, the more questions one asks about it. To get a taste of this, let us do a little experiment. Turn on the camera of your computer and hold a mirror in front of it. Now the camera sees itself. The result should be the same as my looking at myself in a mirror. But it isn't! Something very different happens.

Fig. P-01. My computer camera seeing itself in a circular shaped mirror.
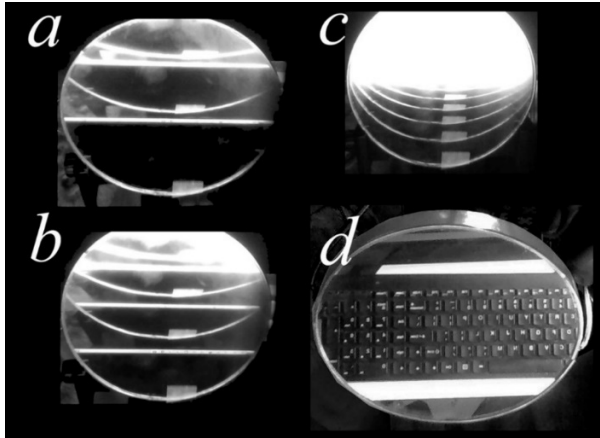
I used a circular shaving mirror for the experiment. Figure P-01 shows the result. At first, the camera produced several staggered images of the rim of the mirror (Fig. P-01a). It appeared as if the mirror was standing in a mirrored cabinet between two parallel mirrors. As I moved the mirror further away from the camera, more and ever brighter rim images formed

that moved closer and closer together (Fig. P-01 b, c). Eventually, they all converged into a blinding bright face of the mirror (Fig. P-01c). Next, the mirror face started to pulsate between blinding bright and pitch black.

All of this happened only if the camera looked at itself. If I turned the mirror and let it look at some other object, e.g., the keyboard, the camera reported everything correctly (Fig. P-01 d).

Here comes the promised disturbing questions. Why doesn't the same thing happen in my brain if I look in the mirror?

The camera lens projects its images onto a piece of silicone with millions of light sensitive pixels on it. The computer reads them and displays pixel-by-pixel an optical copy of the image that the lens saw. True, it's not exactly the same, but the camera can look at it. As a result, the silicone device falls victim to the image of the image of the image… of the mirror. What it experiences is called 'infinite regress'.

In contrast, the human eye suppresses it. But how can it decide which image parts to suppress?

It doesn't. And it doesn't need to, because as soon as the optical image leaves the retina and travels to the visual cortex of the brain, it is no longer an optical image, and will never be one again. The reason is simple. **We never experience anything – consciously or unconsciously – except abstract codes.** After the retinal ganglia have turned the optical image into trains of electrical pulses, **they remain codes from there on**. The eyes will never look at them again as optical images. Hence, there is never a danger of optical infinite regress.

This example demonstrates the advantage of replacing real objects with their codes, but there is so much more to it. Codes are a universe of their own. They are bodyless, weightless, and invisible or transparent, just as the brain finds suitable. No matter how big or complex objects may be, one can add and eliminate them by merely adding or deleting their weightless codes. One can 'look' at objects from different angles and distances by merely re-computing their codes. Codes can be translated, alienated, re-ordered, encrypted, you name it. One can enhance, suppress, or alter the meaning and impact of codes by attaching pre- or post-codewords to them. No miracle is too astounding, not to be performed with codes.

Of course, every computer buff knows all this, and uses it daily in his/her programs. And no software specialist will ever think twice about it.

However, the situation changes dramatically, if we are no longer watching another computer shifting binary data around, but if **we ARE this computer, and the code words are all that we will ever experience of the real world!** Suddenly, the relationship between real objects and their codewords becomes a matter of survival. Now it will be vital to know how

accurate this relationship is, and how much the mentioned code manipulations have altered it. But how can we check them out if we only know codes and never see the real objects?

Evolution has provided four major safety nets for us code-readers in our dark and solitary confinement: (a) **sharp senses, (b) experience, (c) logical thinking, and – as we will argue below – (d) consciousness.**

Being Conscious of Being Conscious

Speaking of infinite regress, the conscious mind can generate quite a few examples on its own. In fact, there are two 'landmines', which can tear human logic to pieces and stand in the way of every attempt to explain consciousness: self-contradiction and infinite regress. They result from the fact that we are conscious of being conscious.

In the first place, this fact funnels all explanations into self-referential statements, which form logical circles, which, in turn, lead back to themselves, and so forth – and we are irrevocably running along an infinite regress – but this time not an optical one. It is a logical one.

However, it also means that explanations must contradict themselves. Assume a person, X, lost consciousness. If every person is normally aware of being aware, then it follows that X cannot be aware of being unconscious. Otherwise, X would be aware of at least that fact, and therefore X would be aware of something. Hence, we must accept that X is unaware of no longer being conscious whenever s/he falls unconscious, and must continue to think that s/he is still conscious (I tested many times whether one can experience losing consciousness by trying to watch myself falling asleep. I failed every time). In short, X feels conscious after having lost consciousness. Try to explain consciousness with this kind of self-demolishing contradiction! The claim of being aware that we are conscious is a big deal. We had better be correct about it.

Whenever I wish to be totally sure that I know something well enough to defend it, I try to defend it against one of my two youngest children. They are – I am naturally tempted to claim – quite intelligent, but more importantly, they are two generations younger. Hence, their way of thinking is a whole Grand Canyon away from mine – which, naturally, they consider to be 'severely antiquated'. In addition, they know everything better than I: If they encounter something new and interesting, they check it instantly on the internet. A few dizzyingly fast finger-clicks on their phone, and they know more details than I will ever know.

Here is my argument that we are aware of being conscious, presented as the transcript of such a dialogue. It is edited, of course, to ensure my victory.

A: Are you aware that you are conscious?

B: No, sometimes I am not conscious.

A: True, but whenever you are, are you then aware that you are conscious?

B: No, I am just conscious.

A: Are you conscious right now?

B: Since you ask, yes.

A: Now try to imagine that you see yourself from a distance, while you are sitting here in your chair. You are conscious, as you have just assured me, and you are consciously looking at yourself. Can you do that?

B: Of course.

A: That means that (a) you are conscious, but (b) you also know that the person you are looking at from afar – namely you yourself – is also conscious. Is that right?

B: Yes.

A: So, you are at the same time conscious and aware that you are conscious.

B: (after a pause) Cool.

Two Biological 'Ghosts': The Temperature of the Body and the Consciousness of the Mind

Whatever explanations of consciousness we may construct, as scientists we think that we must insist on basing them on non-mental, indeed even on inanimate, elements. Otherwise, what good could it do to explain consciousness with elements that are already conscious?

On the other hand, it seems that this is impossible. Consciousness cannot be explained with non-mental, inanimate elements. (Before the reader protests, I hasten to remind him/her that we are talking about a self-contradictory subject!) Let me try to argue this point with a comparison between consciousness and our body temperature.

Both appear like vital, yet weightless, invisible, and indeed ghost-like presences inside us. Both are ubiquitous, familiar, and central experiences of our daily lives. Today, science has brilliant explanations for temperature but none for consciousness. Here are two principal reasons.

1. To measure a phenomenon objectively, we must at least be able to tell its presence. Yet, we do not even know objectively the presence of consciousness in anyone or anything, except inside our own mind. Of course, a conscious person can assure us to be conscious, and we all claim to be able to tell whether a person in front of us is or is not conscious. However, both cases require that the person can communicate with us. If

communication is impossible, there is no objective and independent verification of consciousness. For example, we cannot determine whether, and at what time, a foetus in utero becomes conscious, and we are never sure whether a comatose patient can hear us.

In short, how can one explore a phenomenon whose presence in a study cannot be verified?

2.   Scientific explanations must be testable; or better, according to Karl Popper (1902-1994), they must be falsifiable. Their predictions must be objectively measurable. Therefore, our scientific instruments must be inanimate objects which have no emotions that could distort their data.

Obviously, there are numerous inanimate thermometers that can be hot or cold and thus provide us with their objectively measurable responses to temperature. In contrast, no inanimate object is ever conscious. Hence, observing inanimate objects can never reveal measurable expressions of being conscious. In short, we can easily design and test a 'warm' machine that is built from inanimate elements, but we can never design and test the functions of consciousness in a machine that is composed of inanimate elements!

WAIT! Something doesn't sound right! To identify what it is, let me repeat the last paragraph, while replacing the word 'consciousness' with 'life':

'*Obviously, there are numerous inanimate thermometers that can be hot or cold and thus provide us with their objectively measurable responses to temperature. In contrast, no inanimate object is ever alive. Hence, observing inanimate objects can never reveal measurable expressions of life. In short, we can easily design and test a 'warm' machine that is built from inanimate elements, but we can never design and test the functions of life in a machine that is composed of inanimate elements'.*

This is certainly WRONG. Biochemists and molecular biologists observe non-living elements performing life functions daily. Why, then, did the 'consciousness' version sound reasonable, while the 'life' version is blatantly wrong?

The difference between the two versions is not explicit but hidden between their lines. Look at the 'life' version. Based on the overwhelming evidence for the evolution of life on Earth, biologists have completely accepted the idea that we can explore complex life forms by studying their more 'primitive' predecessors, and we can explore the more primitive life forms by studying their even more primitive predecessors, and so forth, until we finally find it quite acceptable to explore life functions with non-living

elements. (True, we have never created life from basic, inanimate building blocks, although since Friedrich Woehler (1800-1882) synthesized urea from ordinary ammonia, we have synthesized many biological substances from scratch. Even more promising, the students of the so-called 'RNA-world' seem to have come very close to pointing to completely inanimate origins.)

At any rate, general biology has dropped the inexorable demand to explain life exclusively by inanimate elements. Instead, it is making great progress in explaining life based on elements that are still borderline alive, because they may be considered as isolated parts of once living organisms.

Now look at the 'consciousness' version. There is no corresponding general acceptance of studying consciousness in (say) isolated neural tissues, at least not yet. Neither is there a documented evolution of consciousness to justify the permission to use 'not-yet-fully-conscious' predecessors as the elements of explanations of 'more-conscious' successors. To be sure, there is, of course, a known and detailed evolution of the brain anatomy. Unfortunately, that is not sufficient for such studies, as we are not yet sure when and where consciousness resides in any of the intermediary brain anatomies.

In short, we cannot and do not take for granted that consciousness can be explained by isolated parts or studied in 'primitive' predecessors of fully conscious minds. It is for that reason that I stated, in the very beginning of this book, my intention of doing just that.

The Insight of Bertrand Russell

A substantial effort of the present book will be devoted to demonstrating how incredibly elusive and contradictory consciousness is. Part of the reason is, as Bertrand Russell argued, that 'mind' may be more material, and 'matter' more mental, than we usually believe.

However, Russell's insight is also part of the remedy. If he is correct, then there are too many questions about the logic and philosophy of consciousness to be seeking the explanation of consciousness at the level of inanimate atoms and molecules right away. Perhaps it would be better to approach the problem by first looking to 'borderline-conscious', or predecessor forms of consciousness such as 'qualia' and use them as the basic elements for an 'intermediary' level explanation of full consciousness – at least for now.

The reader, who finds this heretical, should be reminded that Russell has been proven right at least in physics and cosmology. Just consider the way these sciences have transformed their concepts of matter by admitting some

strange, immaterial components to their explanations of the material universe.

Of course, the tangible forces and elements of the universe remained still tangible and provable. They simply incorporated the mental phenomena of information, logic and thought. Like metamorphic rocks whose opaque materials are marbled with numerous veins of transparent quartz, matter became marbled with 'veins' of mental products such as invisible logic, thought, and information:

Where physics and cosmology had once relied on gigantic, barely moveable masses and irresistible forces of gravity, there is now the abstract and immaterial space-time and its contorting curvature. Where once there had been indestructible particles, there are now ghost-like, empty quantum fields and their abstract symmetries. Obviously, curvatures, empty fields, and symmetries are not the defining properties of hard matter, or are they?

In retrospect, it is also obvious that immaterial components had been part of all our explanations from the very beginning of the exact sciences. Indeed, we had always needed to invoke the immaterial phenomena of thought, logic, and information. Like the ancient gods we once worshipped, those three mental phenomena are invisible and intangible. Like those gods, they are unprovable. Instead, their existence relies on axioms.

Hence, something powerful and mysterious that is comparable to the invisible ancient gods of our past has gradually crept back into our exact and impersonal scientific explanations. And, as suggested by the multi-billion-dollar profits of the internet industries which sell thought, logic, and information, we are still worshipping them.

What To Do About Unorthodox Explanations

Alas, by considering such 'mixed breed' elements to explain consciousness, the resulting explanations would no longer seem as 'clean' as the earlier explanations that were based exclusively on non-mental elements. Instead, they could appear to be somewhat 'murky' because they seem to cheat.

But the 'clean' elements of explanation have not got us very far. To date, we have no explanations of consciousness that are based exclusively on non-mental elements. Indeed, we have no explanations of consciousness at all. So, we might as well try out the somewhat 'dirty' explanations and 'clean them up' later.

It reminds me of Werner Heisenberg (1901-1976) and Carl Friedrich von Weizaecker (1912-2007) doing the dishes after their dinner in a ski hut, when von Weizaecker asked the rather profound question: 'do the

philosophers understand how dirty water and dirty towels can get dirty dishes clean?'

I daresay that nobody knows how we clean up our misconceptions. Nevertheless, we often trust that it works. **Dirty water and dirty towels will eventually get the dirty dishes clean enough to reveal some hidden patterns in their glaze**.

CHAPTER ONE

THE ELUSIVE, CONTRADICTORY, ENIGMATIC, FASCINATING NUISANCE OF CONSCIOUSNESS

Judging by the large number of books, internet talks, and websites that discuss consciousness, one should think that people love the subject. Yet, consciousness is not even a likable subject. In fact, there are many reasons to dislike it intensely. Already, the Bible considers the desire to be conscious as a sin:

> *'…And the woman said unto the serpent, We may eat of the fruit of the trees of the garden: But of the fruit of the tree which is in the midst of the garden, God hath said, Ye shall not eat of it, neither shall ye touch it, lest ye die. And the serpent said unto the woman, Ye shall not surely die: For God doth know that in the day ye eat thereof, then your eyes shall be opened, and ye shall be as gods, knowing good and evil…'* (Genesis 3).

So, Adam and Eve ate the forbidden fruit. The punishment for the sin of wanting to be conscious and to 'know good and evil' was swift and harsh. Adam and Eve, and with them all of humankind, were expelled from the Garden of Eden forever.

Neither general language, nor modern jurisprudence has a better opinion of consciousness. For example, we describe a criminal person's character as 'inhuman' only if the person's evil actions are conscious. Correspondingly, consciousness is linked legally with the evil of criminal intent: no act can be evil, and thus punishable, unless it is committed consciously.

Medicine considers consciousness as the enemy, too. Obviously, anaesthesia is considered as a major blessing of medicine, but what else is anaesthesia than the local or global temporal suppression of consciousness?

Turning to history does not offer much encouragement, either. In fact, history tells of horrifying weapons that were invented and applied by consciousness. Would humans have invented them if they had remained permanently subconscious organisms? Worse, every conceivable human atrocity such as wars, genocide, holocausts, race discrimination, slavery, euthanasia, experimentation on human subjects, and many more were

'justified' and blessed using reasons concocted by the conscious human mind.

On the other hand, consciousness is undoubtedly a marvel of biological engineering. As an accomplishment at the top of the evolutionary chain, one should rightfully expect that it is a powerful instrument of survival.

Well, that seems hardly true, either. Worse, it is often an outright handicap and hazard, especially when a fight for survival must be fast, sure-footed, and daring. In critical situations where we cannot afford any mistakes, we usually switch to instinctive, subconscious controls.

But even when not fighting for our individual survival, we stumble badly whenever we try to inject consciousness into fast-paced physical and mental activities. My present typing on a keyboard will slow to a standstill if I try to aim consciously at the various letters and send my fingers to the corresponding keys. Or try to play a musical instrument consciously! It's not possible. Highly accomplished pianists usually play with closed eyes, in a trance.

Since it does not help the survival of the individual, is consciousness at least a delightful luxury?

Absolutely not! We distrust and shun consciousness whenever we can. In fact, we seek happiness and even ecstasy by turning it off. Gigantic industries have been created and flourish providing sexual illusions, thrills, and drugs to help in 'numbing' the mind by silencing consciousness.

Can consciousness at least borrow some usefulness from its close relative, the 'attention'? Obviously, attention is a vital ability. On a planet where predators hunt prey, and prey must escape predators, it is self-evident that the ability to notice and focus on every detail of the environment has a vital survival value. But what about consciousness, which may be viewed as some form of 'self-attention'? Would it be equally beneficial for survival to pay as much attention to oneself?

Definitely not! Ask any martial arts practitioner! Focusing on yourself instead of on your opponent is a certain recipe for your swift defeat.

I.1. The Inevitable Self-Requirement

We are conscious of being conscious. So, consciousness is conscious of itself. Inevitably, therefore, all attempts to explain consciousness must run into the situation which was famously depicted by Maurits Cornelis Escher (1889-1972) in his lithograph of a hand drawing another hand which, in turn, draws the first.
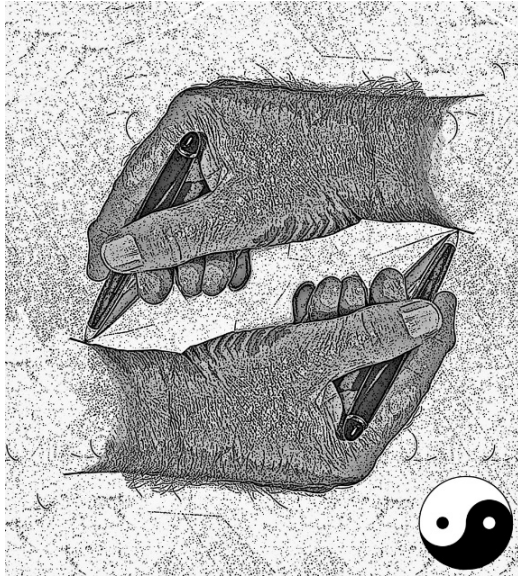
Fig. I-01. Self-reference. My right hand drawing itself, reminiscent of the ancient Chinese idea of Yin and Yang.

Adopting Escher's idea, Fig. I-01 shows my right hand drawing itself. It is not only a troubling circular thought, but a serious obstacle to the logical analysis of consciousness. But wait! The idea of something creating itself – or its opposite – is much older than Escher. It is at least as old as the ancient Chinese idea of Yin and Yang, which carries the idea even further as the two mutually created objects may combine into a single object of perfection, namely a circle (lower right corner of Fig. I-01). Far from being interpreted as an obstacle to logic and consciousness, it is considered as one of the deepest philosophical concepts.

So, is self-reference good or bad? Before answering the question, let me explain some of the illustrations which I use for the answers.

I.2. Basic Model and Its Illustrations

All discussions of consciousness need to introduce a person who is conscious of something, preferably but not exclusively, of the outside world.
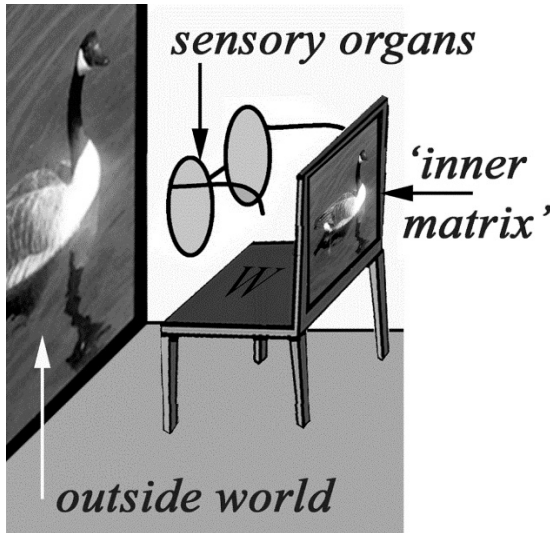
Fig. I-02. The graphic representation of a 'watcher' as a chair with glasses who watches a real-world scene (a Canadian goose) which is projected to its inner matrix.

We will call this conscious person, a 'watcher', and depict all watchers as a chair marked with a 'W' on the seat (Fig. I-02). The sensory organs of the watcher are symbolized by a pair of glasses that look at a scene in the outside world. The sensory organs project the outside world onto a dedicated part of the brain, which will be called the 'inner matrix'. The inner matrix of the watcher is represented by the image of the outside world projected onto the back of the chair. In most cases, the outside world is depicted as a group of Canadian geese and their goslings swimming in a pond.[2]

---

[2] The basic idea of symbolizing a conscious watcher with someone observing a scene in the outside world displayed on a 'stage' is at least 2,300 years old. In his dialogue 'Republic', Plato (appr. 428-347 BCE) compares us with captives in a cave. We never see the real world except the shadows of things cast on the cave wall in front of us by a fire at our back. Modern philosophy has adapted and augmented the metaphor. Today's philosophers and psychologists use other metaphors for our inability to see the true reality. As pointed out by Bernard Baars in *A Cognitive Theory of Consciousness* (Cambridge University Press 1988) the modern versions of the cave metaphor instead use the metaphor of a 'Theater of Consciousness' with metaphoric audiences, actors, directors, stagehands, and so forth.

The watcher, following an instinctive compulsion, reads the 'images' on the inner matrix, and – as will be described later in more detail – converts them into 'qualia' which represent the personalized experiences of specific objects and their colours, shapes, sizes, sounds, etc., as displayed on the inner matrix. Since they are personalized experiences that cannot be observed from outside the mind, the illustrations label them as 'classified' (Fig. I-03). Each entry in the inner matrix gives rise to multiple qualia.
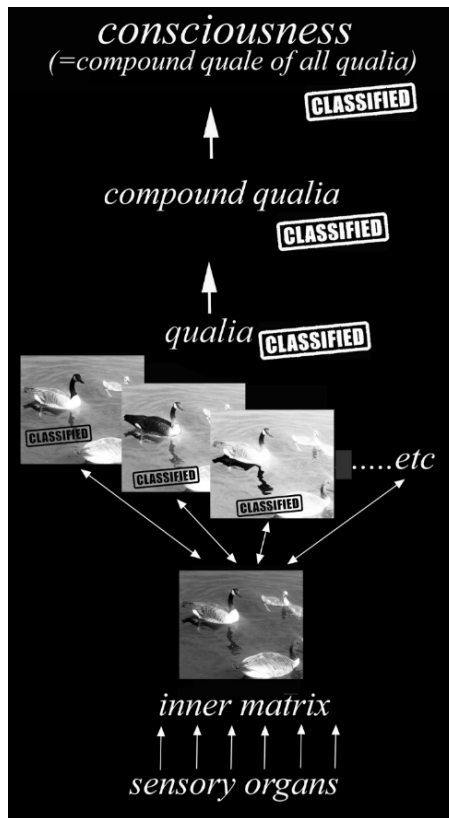
Fig. I-03. The graphic representation of the main hypothesis of the present book. After the sensory organs of a watcher have projected the real world onto the inner matrix, the watcher's compulsive reading of the inner matrix leads to the creation of the 'qualia' (labelled as 'classified'). Next the qualia are compacted into 'compound qualia' of increasing numbers of components. The compound quale of all present qualia is, then, the quale of consciousness.

The various qualia do not only project a specific object and its properties in isolation, but each quale also integrates the object's context/environment. Like the object, its context is read off the inner matrix as symbolized by the double-headed arrows in Fig. I-03. As discussed in chapter III, the inner matrix may store and recycle past images. Hence, part of the context of the objects may be their past.

The next step in the processing of the qualia is the combination of several 'prime' qualia into a 'compound' quale, which represent composite objects.

Some of them contain a handful of components; others contain millions. One unique compound quale stands out among all others, namely the one that unites all qualia that are presently in the mind. Like all other qualia, it personalizes the experience of its collective components. In the sense of the 'running' average of a set of changing numbers, it is constantly updated by incorporating every new quale that appears in our mind. Therefore, **let me describe it as the largest 'running compound quale'. This very quale will be interpreted as consciousness** (Fig. I-03). (The reader, who worries how such information-rich qualia can remain small, will have to wait for an explanation until chapter III).

I.3. Explaining 'Explanations'

Before explaining consciousness, one must decide what qualifies as an 'explanation'. The word 'explanation' and its synonym 'explication' are both rooted in Latin. They are composites of 'ex' (out of, from) and either 'planus' (flat) or 'plicare' (bend, fold, braid). The literal translations do not seem to be quite on target. 'Explanation' seems to describe what one finds inside a convoluted object after flattening it out. 'Explication' seems to describe what one finds inside of its folds and braids.

Obviously, there are also phenomena whose 'revealing of their innards', would only increase the bewilderment. For an example, just consider the cockpit of a modern airliner!

The field of formal logic has a much more accurate definition: **every explanation is a tautology.** The word 'tautology' is derived from the Greek 'tauto' (same) and 'logos' (reason, word). Hence, it defines explanations as statements that formulate the same subject in two different ways. The first way names the subject in common language, while the second way must describe the same subject in the words of its fundamental 'explanation-language'.

Take the example of the explanation of a 'crime'. According to the Oxford Dictionary, crime is 'an action or omission that constitutes an

offence that may be prosecuted by the state and is punishable by law.' It is a typical tautology, as it formulates the concept of 'crime' on its left side in common language, and on the right side in the 'explanation-language' of jurisprudence.

However, it takes a lot more than finding an explanation-language to produce an explanation. Every explanation must fulfil the following four requirements:

1. A declaration of axioms (unexplained postulates and phenomena) from which it will derive the explanation (**explanation-language**).
2. A set of observed expressions of the subject that it aims to explain (**target phenomena**) in this language.
3. A string of logical conclusions that apply the axioms and bring about those observed expressions (**hypothesis of the target phenomena**).
4. Several tests to verify the explanation (**validity tests for the hypothesis**).

I.3.1. The Dragnet of Necessary Conditions

Whenever science seeks explanations of a phenomenon, it begins by establishing a list of its 'necessary conditions'. Conditions are necessary if the phenomenon cannot happen without them. For example, wings are necessary for birds, because they cannot fly without them.

In general, the necessary conditions are not sufficient. For example, still owning its wings would not make a dead bird fly. Inevitably, lots of additional necessary requirements must also be fulfilled. For example, aerofoil-shaped wings, hollow bones, air sacs, a fast-pumping and heat-maintaining heart, a cloaca that minimizes the water content and weight of a bird's body, an aerodynamic body shape, turbulence suppressing feathers, and many more conditions must be fulfilled before a bird flies.

Naturally, everybody feels confident that increasing the numbers of necessary conditions must eventually lead to a sufficient explanation. That is the reason that many scientists in search of an explanation keep identifying more and more necessary conditions for their research topics.

However, it is not the numbers of necessary conditions that matter, but their weight. For example, assume you are an alien scientist, who tries to explain human sewing machines! You could amass thousands of necessary conditions that, nevertheless, would apply as much to sewing machines as to bicycles. You would not even come close to an explanation of the workings of a sewing machine, until you find its most pivotal

property: sewing machines must have needles, whose eye is in the tip. Only this necessary property would get you close to an explanation.

Usually, we have no proof that a list of necessary conditions is also sufficient until we can make the phenomenon happen with our own hands. For example, the sufficiency of Newton's laws for celestial mechanics was only definitely proven after we had launched our own satellites circling the Earth.

Accordingly, we will not know whether a list of necessary conditions for consciousness is sufficient until we have created artificial consciousness. Until then, we must continue to collect the necessary conditions of consciousness.

So, how would the hunt for consciousness work? What are the relevant necessary conditions of consciousness?

Apart from the obvious requirements of a functional cardiovascular system and other basic requirements of the living state, most other necessary conditions of consciousness are the **absence of the traumata** that are known to cause the loss of consciousness. Examples are the absence of cerebral haemorrhage, dementia (such as Alzheimer's disease), head trauma, brain tumours, heart disease, heat stroke, liver disease, uraemia, seizures, shock, drugs, alcohol, low blood sugar, or lack of oxygen to the brain.

Most of these conditions are little help in pinpointing more specifically the brain loci that must be fully functional for consciousness to operate.

I.3.2. The Search for an Explanation-Language

According to the above formalism, every explanation should begin by defining a suitable explanation-language. What is the right explanation-language for consciousness? It should define many more words for related details and actions than the common language. It should have a large range of applicability, i.e., it should apply not only to matters of consciousness, but to other mental products. Its vocabulary should be complete, consistent, and unambiguous. Its rules of grammar should reflect fundamental interdependencies operating in the mind and brain.

Do we have such an explanation-language? Not yet, although there are very promising, indeed brilliant candidates, such as the languages of philosophy, psychology, neuroanatomy, neurophysiology, electrophysiology, cell biology, biochemistry, molecular biology, genetics, quantum mechanics, and the mathematics of iterative, i.e., chaotic systems. Each one of them offers powerful, provable explanations for several aspects of consciousness. However, none of them covers all aspects. Time and, again, new thoughts

and experiments demand the updating or revision of their vocabulary, rationale, and rules of linking their terms.

Another major problem of explanation-languages is their eternal battle between doing justice to details on the one hand and facilitating our understanding on the other. This has to do with the fact that all real things are fractal in nature. The detailed aspect of consciousness forces the explanation-language to offer explanations for the words it uses. Subsequently, it needs to explain the explanation of the words, and again to explain the words used by the explanation for the words, and so forth. This unstoppable process leads to an ever-greater fragmentation and specialization of its subjects, which in turn drowns it in *ad hoc* invented, unfamiliar words, and moves the understanding of the subjects into ever greater distances.

I.3.3. How Do We Know That We Have Found an Explanation?

I believe it was the military theorist Carl von Clausewitz (1780-1831) who urged all military leaders to, "before starting a war, define victory". Especially recent history offers several disastrous examples, where warring nations ignored this advice, and were forced to forever continue smouldering conflicts.

The same applies to intellectual wars. Therefore, it may be wise to follow Clausewitz's advice and define 'victory' by deciding at what point we may consider consciousness as explained.

Considering the enormous sophistication and variability of consciousness, I would not expect the explanation of consciousness to be a small and aesthetically pleasing set of equations. Instead, it is more likely to be an elaborate computer simulation with algorithms that make the computer appear to be conscious, as determined by something like a modified Turing test. But which kind of modified Turing test should it use?

A long time ago, I thought I had found a test for computer consciousness. I had just saved enough money to buy my first personal computer, a Radio Shack TRS-80 Model I with a Z-80 microprocessor and a RAM memory of 4KB (yes, only puny 4 kilo-bytes, which seemed fantastically large at the time).

It was a wondrous thing. Soon after I had learned to program it in machine language, a simple question popped into my mind. 'Can I write a program that makes my computer scared?' Of course, even beginners could let a computer print the words "I am scared", but I meant something very different. I wanted the computer to 'feel' scared, and not just act scared - freezing up, stammering, sweating, trembling, babbling, crying, trying to run, whistling an old computer song like HAL…

Many faster and more powerful computers and several more sophisticated compiler languages later, I am still asking the same question, and I still do not know the answer.

Let us not forget the obvious reality. No matter how fast or powerful they may be, all of today's computers do no more than transform one set of symbols into another set of symbols. They never turn the symbols into conscious experiences for themselves. Perhaps quantum computers can change that. Perhaps, the qubits of quantum computers can include scared states among their multiple states?

Assume we manage to construct a computer algorithm that generates the symptoms of fear inside itself. To learn the secrets of consciousness, could we not simply monitor its every single transistor? Would it put us on the track of understanding consciousness if we find something spooky happening in the circuitry?

Hardly! To simulate consciousness, we want the computer not only to feel scared, but also to know that it is scared. How would we know what it knows? We cannot know that for the same reason that we cannot know the conscious mind of anyone or anything other than us.

Therefore, I am afraid that we must drop the hope of producing a silicone-based simulation of consciousness and turn to our own mind and its mental products if we want to convince us that we have found an explanation of consciousness: no matter how paradoxical it sounds, **only our conscious mind can decide whether we have found an explanation for consciousness**. We must convince it that its secrecy has finally been defeated, and the explanation has been found. **We will not know that the explanation has been found until we feel the same certainty that we feel about the proofs in Euclidian geometry.**

I.3.4. Which Face of Consciousness Should We Explain?

We can all agree that consciousness is a 'something or another' of the mind. Is it…?

1. **…a state** of the mind?
   ("The state of consciousness is suspended during sleep.")
2. **…an experience** of the mind?
   ("The clarity with which consciousness filled K.'s mind, dispelled all fears…")
3. **…an ability** of the mind?
   ("Only the conscious mind can understand mathematics.")

4. **…a content** of the mind?
    ("The memory of her first conscious hearing of Bach's 'dona nobis pacem' brought tears to her eyes…")

Before we decide which of the above cases apply to consciousness, we must also decide which mind we are talking about.

I.3.5. Which One of Our Minds' Consciousnesses Should We Explore?

The above list of faces of consciousness assumes that we have a single mind. Our reason is as obvious as it is naive: we experience only a single consciousness, which in turn keeps us aware of only a single 'me'. Yet, in the past, everybody who tried to define 'mind' as the source of our cognition and wilful actions had to accept that there are at least two minds, the 'conscious' and the 'subconscious'. The 'conscious' mind asks and formulates the questions about 'mind' in rational terms. The 'subconscious' mind is the source of our irrational compulsions, which includes our curiosity about 'mind' in the first place.

In general, we lump the phenomena that are linked to language, logic, and planning into the conscious part of the mind because they permit observation, description, classification, experimentation, and analysis. Everything else, like faith, courage, love, creativity, conscience, and many others, remains in the subconscious mind.

Inevitably, then, the conscious mind will appear as the shining, credible, logical, and invulnerable one of the two, but it must also appear as the more superficial one. For example, if Vincent van Gogh dips his brush into oil paint, we generally consider it as a conscious but unimpressive act. By contrast, we credit van Gogh's subconscious mind with the far more profound act of choosing the perfect colour, shape, and place for his next brush stroke on the canvas of the 'Irises'.

In this sense, the subconscious mind appears as the source and home of all the unfathomable and mysterious mental products that motivate our actions such as energy, mystery, will, emotions, ethics, courage, endurance, religiosity, and creativity, but, alas, also everything that is evil and destructive.

Based on that distinction, it sounds silly to ask which consciousness we should explore, the consciousness of the conscious or of the subconscious mind? After all, by definition, the subconscious mind does not have a consciousness. What is there to explore?

Analogous to Russell's insight into matter and mind, the two minds are not entirely separate, rather their actions are linked and depend on each other. We must be prepared to observe 'feelings' of consciousness, which should belong to the subconscious mind, but also 'logical relationships' between emotions, which should belong to the conscious mind.

This distinction between the two minds raises the question, how do we know that there are only two minds? Couldn't it be that some of us have only a 'standard pair', while others have many more consciousnesses? Of course, the answer is that we are not aware of any others. But that is a circular argument. If other subconscious minds were to exist, we would include them in the subconscious mind. If there were other minds that are conscious of the same body and the same universe, we would consider them as the minds of another person. Indeed, patients with multiple-personality syndrome seem to have only a single consciousness at a time.

Yet occasionally, our seemingly singular conscious mind reveals almost alien new faces. Even perfectly normal people seem to occasionally act as if they had multiple personalities. How else could writers and actors slip into the multiple characters they describe or enact? Also, powerful drives such as sexual desire, rage, or fear seem to bring character traits to the surface that surprises not only bystanders and friends, but the person him/herself. It is not clear whether these changes of character are the result of switches to alternate instances of consciousness, or merely expressions of altered states of a single consciousness.

In short, we do not know how many minds each of us has, and how many minds we must explain if we seek to explain consciousness. Hence, we cannot choose one of the faces of consciousness as the subject of our intended explanation, unless we decide which mind's face it is.

But let us postpone that decision as well, until we have answered yet another question: whose mind we are talking about?

I.3.6. Whose Consciousness Should We Observe?

A most important question in the experimental sciences concerns the choice of a suitable test system. In our case, we must decide whose consciousness we should study. Considering that we all experience consciousness as a natural state of our minds, the experimenter might consider studying consciousness inside him/herself. On second thought, how can looking into one's subjective experiences and feelings be a sufficiently objective measure to satisfy the minimal requirements of science?