

[Open Peer Review on Qeios](#)

# Doyle's Two-Stage Approach to Free Will: A Biophysics for Real Choice?

Jonathan Edwards<sup>1</sup>

<sup>1</sup> University College London, University of London

**Funding:** No specific funding was received for this work.

**Potential competing interests:** No potential competing interests to declare.

## Abstract

Robert Doyle has argued that traditional difficulties associated with the concept of free will can be resolved by a two-stage approach, explored by several philosophers and scientists. Possibilities for action can be generated within the brain, more or less randomly, from which an 'adequately determined' choice can be made. A similar selective process is well worked out within immune cells, which may provide a useful model. A degree of randomness is pervasive in the physical world but living systems have evolved to both tame and exploit this aspect at the interface between fluid and solid phases. Random and systematically determined processes are used in combination. These cell biological considerations support the plausibility of the two-stage model and may help point to specific mechanisms but may also raise a question about 'who' is doing the choosing. Computer modelling may sharpen the focus of that question. It may also highlight a paradox – that 'freedom' may only exist as the flipside of limitations or constraints imposed by real world situations.

## Introduction

Robert Doyle (2013) has made the case for a two-stage model for free will in which a degree of randomness in events in the brain generates a range of possibilities (or representations of possibilities), from which an 'adequately determinate' selection can be made. This provides a basis for the notion of free will, or choice, but with the 'freedom' perhaps dissociated from the will or choice.

Versions of this model have been arrived at by several philosophers and scientists from William James (1890) and Henri Poincaré (1958) through Daniel Dennett (2003) to Martin Heisenberg (2009). Despite the range of names associated with a two-stage model, several have raised uncertainties about how it might work in practice or even whether it is plausible. Proponents have differed in their attitudes to indeterminacy, its relevance to macroscopic events and whether the model still requires some form of dualism to meaningfully provide 'freedom'. The aim here is to try to address uncertainties with some practical proposals for a two-stage form of free choice in the context of neurobiology. The implications for free will in man-made computer systems (Blum and Blum, 2022) will also be touched on.

Several salient questions arise in Doyle's discussion of two-stage models.

1. Is there genuine indeterminacy in the world?
2. Can indeterminacy at a fundamental (quantum) event level show through at a biological scale in a way that could matter?
3. How could random events be programmed in to brain function?
4. Must all indeterminacy be random, as in Hume's claim that events are either determined or random, with no intermediate option, or does creativity, or novelty, involve a third way?
5. What would an 'adequately determined choice' be – another name for a recognisable physical process or something other?

In a nutshell, I shall propose that answers to these questions depend on the way modern physics tells us chemical reactions behave in the part-fluid/part-solid microenvironment of living cells, and how cells exploit this to survive. The computer programmer exploits a very different environment in a different way, but the same basic truths may show through. Every seeming solution to the problem of free will threatens to reveal further layers of inconsistency, and some will be touched on, but only so much can be covered at a time.

## Question 1. How does modern physics provide indeterminacy?

The focus of this essay is on the way known biological mechanisms, particularly in the immune system, can support and clarify a model of free will. Nevertheless, indeterminacy of those mechanisms derives from basic rules at the molecular, i.e., quantum, level. Some initial discussion of how, is relevant. Chemistry is all quantum electrodynamics, so biochemistry is too. That might sound obscure, but the origin of randomness in quantum theory is simpler than often suggested.

A lot of discussion of quantum-level indeterminacy focuses on measurements of 'observables' in a physics lab – properties like spin - attributed to individual field quanta, often loosely called particles (or wavicles). What may get forgotten is that events in physics labs are designed to check basic theories, and that involves making use of complicated rules about what information one can extract under tightly controlled conditions. To make things intelligible use is often made of the idea that a 'wavicle' progresses in a 'superposed state' according to deterministic rules and *then* randomly opts to collapse into one possible final state. Yet a fundamental tenet of the theory is that a field quantum is a single indivisible causal connection between a set of 'initial conditions' and some 'final conditions'. There are no *actual* states in the usual sense of the term, in between. There is, therefore, no reason to think wave function collapse is anything more than the transition from an account of what events might happen to an account of which one did happen. Moreover, there is no need to see randomness as a mysterious, unexpected property of quantum theoretic formalism. There is a simpler way of seeing why randomness has to be part of physics.

Indeterminacy in fundamental quantum events arises from three basic factors: event individuality, *symmetries* and Heisenberg's Uncertainty (or better Indeterminacy) Principle. Events involve individual quanta that can only happen this way or that, not both. If there are several symmetrically suitable options, randomness is hard to avoid. An example of symmetry is when a photon is emitted, as an electron falls to a lower energy state. There are an infinite number of

directions to choose to go in. The chance of going left is  $p=0.5$ , and right 0.5, because no rule says which should be preferred. An interesting chemical symmetry is chirality. If a mixed solution of molecules, none of which are big enough to have a left- or a right-handed ‘elbow’, sets up a reaction that generates a molecule type with an elbow that must be left- or right-handed, you get equal (symmetrical) numbers of each. For any individual event the chance of a left-handed elbow is  $p=0.5$ . Chirality was discovered by Pasteur; long before quantum theory, chemistry showed its indeterminacy.

We might argue that, for every interaction at small enough scale, we can find asymmetry. A billiard ball might always hit another slightly to right or left. An electron might be swerving round a nucleus just this way, so send a photon just there. This is where Heisenberg’s Principle comes in, providing an empirical validation of an *a priori* requirement of Leibniz’s Law of Continuity (*pace* Laplace), that there can be no fact of the matter below a certain scale exactly where a dynamic interaction occurs (Edwards, 2023). The only way out, for ‘God’, is, rather than to reverse probabilities from  $p=1$  for left and  $p=0$  for right over an *infinitesimal* distance, to make a gradual shift via  $p=0.5/p=0.5$ . Even where the line-up is well off, we do not quite get  $p=1$  and  $p=0$ . Everything is slightly indeterminate. The simple way of explaining this is that interacting quanta involve notional oscillations, which must have wavelengths, and to have a wavelength something must spread out over at least a wavelength or two. This ‘wavelength’ may just be an epistemic trick to help us picture the rules, but those are what we are looking for, so it does the job.

It might still be argued that all the randomness at the molecular level can be dealt with by invoking wave function collapse, but it cannot, because that tells us nothing about the rate at which quanta of a certain type *come into existence*. It fails to explain why a good claret takes years to mellow in bottle. The reverse probably *is* true – we need no concept of collapse. Rather than suggesting that a quantum with partial identity somehow ‘on its way’ jumps into the identity of one of a range of complete quanta, it is much easier to suggest that the equations just indicate that the probabilities of each of a range of fully identifiable (complete) quanta form a ‘wavy’ (i.e., periodic) spectrum, some being much the same and others more, or less.

Another factor that physics lab experiments tend to bypass is that once a particular quantum-level event has happened in a biochemical soup in a cell, the options for the next event change to a new set of possibilities. One might argue that in fluid phase, where there are billions of copies of each molecule nearby, one event does not change things much. However, as I shall come to, this is not how biological systems do things – or at least not some things.

The reality of life inside an oak leaf cell is that at any point in time an infinite range of quantum events might happen next. All biology is essentially photons interacting with electrons, and orbitals swapping around. If, for instance, a blue photon arrives, it may be absorbed by an electron in chlorophyll, or it may not, or it may be absorbed or reflected by a vast array of other cell components. The minimum interaction domain of a photon will overlap with millions of potential electron partners. If the chlorophyll electron passes to a new state, that may decay with release of a new photon after 0.73 picoseconds or 2.40 picoseconds or ... whatever. And so on.

Neither the biologist, nor, as far as I know, the cell, is interested in what spin a photon has. They are interested in *which* of a potentially infinite number of photons *happens*, in the sense of connecting some prior events to some outcome,

something of relevance to Question 4, below. The physicists' checked theories predict all sorts of quantum events as more or less likely to occur in a given situation. Born's (1926) rule, that the equations tell us about probabilities, ultimately translates into this. But it does so through generalisations that leave complex numbers behind - for instance, that an infra-red photon will almost certainly not excite the chlorophyll electron.

Some events, like two electrons sharing an orbital with the same spin, are banned, but for all permitted events (defined by an arbitrary window of options to get a finite value) there is always a certain probability of occurrence, of *somewhere between 0 and 1*. A degree of randomness pervades physics, for very simple reasons.

## Question 2: Can fundamental-level indeterminacy show through at biological scale?

The fundamental reason for randomness is that rules of interaction involve symmetrical possibilities but are individual events. All situations have several legitimate outcomes. If we define any next event in a biological context precisely enough – down to one specific change in one molecule – its probability of occurrence at a time point will be tiny and a vast number of other events could occur instead. Yet life has developed ways both to tame this randomness and to make use of it.

In a liquid solution in which zillions of chemically similar events can occur, as in a test tube, the probabilistic nature of individual interactions might become irrelevant. It does show through, as in the example of an equal racemic mixture of left and right-handed molecules appearing from achiral precursors. Yet without knowing Heisenberg's Principle we could argue that strict determinism could give the same result. Solid objects tend to behave as a few individuals, as in a hammer hitting a nail, so there is no evening out by numbers. On the other hand, we have reason to think that the behaviour of these solid objects is explained by vast numbers of atoms working together in a way that can be predicted well enough by classical physics with  $p=1$  for hitting the nail and  $p=0$  for missing.

Life, however, bucks the system. Inside a cell some molecules are present in billions, but others in very small numbers. Most notably, the molecular structures of genes are only present in two, and often in operational terms, just one, copy. A key feature of these 'rare' molecular structures is that they tend to involve both repeating units and marked internal asymmetries in terms of limited numbers of units or variable unit type. Schrodinger predicted these supermolecules in *What is Life* (1944), suggesting that 'the heredity material is likely to be a molecule, which unlike a crystal does not repeat itself'. Cytoskeletal assemblies supporting domains like neuronal dendrites are another example of molecular-level asymmetry that straddles fluid and solid phase dynamics.

Interactions of genes are highly predictable because of their asymmetry. Mostly they bind in one orientation to messenger RNA or histones that fit like lock and key. But during reproduction germ cells undergo meiosis in which the genes have a masked ball. The pairs of chromosomes separate into sets, then enzymes cut between genes and swap strings of genes from one chromosome onto its pair – chromatid exchange. The options are still constrained but there is freedom in that each enzyme molecule can cut at several points (see Verhest and Heimann, 2008). This is a plausible source of gene reduplication. Asymmetric exchange leaves one chromosome with two copies of a gene, which can evolve independently

to serve two functions. At every stage a single molecular or quantum level event is involved. A domain of the antibody (immunoglobulin) gene turns up in hundreds of proteins, possibly through this mechanism. In this way molecule-based (quantum level) events can not only change the behaviour of millions of cells in a resulting embryo but have a dramatic influence on the planet for millions of years by giving rise to successful new species such as pampas grass. Richness of life on earth is all due to indeterminate quantum-level events showing through at organism and species scales. These are not chance *mistakes*. Their occurrence is programmed in to the genetic process.

What may be less well known is that this occasional tendency for genes to engage in indeterminate interaction is part of the maturation and survival of an individual human. We know how important it is in the immune system and there are good reasons to think something analogous is at least as important in the brain.

The function of a B lymphocyte is based on one molecular species – the antibody it produces. Every day we produce a billion B lymphocytes, and each makes a different antibody. There aren't a billion genes to do that. Moreover, some of these cells try making one antibody and, finding it no good, switch to a new one, and those whose antibody species turns out to bind usefully to microbes can shift to making slightly different antibodies again, that may bind better.

All of this is done by enzymes, including activation induced deaminase, whose role was worked out by Michael Neuberger (2008). As in meiosis, the enzymes cut genes and shuffle them into new positions. In addition, they can take out a nucleotide and *randomly* replace it with one of the other three options. As a result, the antibody protein encoded by the gene is a bit different for each cell. The advantage is that the immune system has antibodies available to bind to millions of possible microbial proteins. There is no way of making an antibody by *instruction* using the microbe protein as template, so the system relies on making a vast number of antibodies to cover *all possibilities* and *select or choose* a useful antibody having found it to bind to a target. Cells committed to a chosen antibody species are allowed to multiply and produce antibody on industrial scale as plasma cells.

The implication of all this is that to have a good chance of surviving to adult life we rely on random individual molecular or quantum level events that can show through as large-scale protein synthesis. The events are not perfectly random, but they are close to it, perhaps unsurprisingly since the advantage of the mechanism is to generate as wide diversity as possible. Interestingly, nucleotide changes seem to be partly '*pseudorandom*' in that the randomness is partly contrived by enzyme systems ensuring that alternatives that might not, by default, get an equal crack of the whip, pretty much do so.

In James's (1890) two-stage free will model one component of the randomness that generates possibilities is attributed to chance encounters with the environment. That component places the 'free' aspect of the model outside the individual. What we now know about programmed internal randomness confirms James's claim of another, internal, component that can make the free aspect the responsibility of the individual – at least in that it resides nowhere else. That is relevant to arguments about morality and justice, at least in pragmatic terms.

At the other end of the spectrum, Popper (Eccles and Popper, 1984), and others since, have suggested that the random possibilities exist at the level of 'superposed' quantum states, with 'choice' involving the restoration of determinacy with 'measurement'. However, this is *not* how the immune system uses randomness, which requires no such speculative

interpretation of quantum theory. As indicated above, the concept of 'superposed states' may only generate unnecessary confusion.

Doyle, himself, sees randomness as inherent to biological processes, much as suggested here. He uses the term 'noise'. While not unreasonable, that may perhaps oversimplify a mechanism we can analyse in detail both at fundamental physics and molecular biological levels. Biology harnesses randomness in a very precise way.

### Question 3: How to set up possibilities for choices in brains

The situation in the brain is less well worked out. However, the B lymphocyte story provides an obvious template for a two-stage model of free will. It involves the generation of a vast range of options and the choice of the most useful. Moreover, there are likely parallels with brain development and plasticity. The general concept was championed by Edelman in his *Neural Darwinism* (1987).

Random generation of antibody allows for formation of memory banks of cells that recognise targets and can be selected for expansion if a target is re-encountered. The daily production of a billion options maximises choice of fit for new targets. If chosen, these can undergo plasticity to generate more exact matches.

All these aspects may be relevant to brain. Neither B lymphocytes nor neurons have mechanisms for creating recognition patterns purely by instruction. The chemical engineering necessary is not within the scope of living cells. For lymphocytes it would require not only a reverse transcriptase like that used by retroviruses to insert instructions into DNA, but also 'reverse ribosomes' to back-translate protein sequences, not to mention a way to generate 'lock for key' proteins systematically.

For the brain, some element of instruction occurs. Patterns of neuronal firing can be captured from a single fleeting memorable event by synaptic reinforcement. That can give possibilities of the first sort James proposed. However, it is not so easy to explain how each event or image can be selectively captured and stored such that it can be retrieved. A mechanism for pre-generation of varied cell response repertoire is likely to be needed, with the closest fitting cells volunteering to take a new memory, with options to fine tune later. As for lymphocytes it would benefit from having lots of cells. Moreover, for simultaneously representing many possible *future* options for actions and predicted consequences there is no obvious mechanism for instruction.

To give a bit more detail; each brain neuron sends a branch of its output axon to each of about 10,000 other neurons. A cortical pyramidal neuron likely feeds cells of several functional classes but may send at least half its branches to neurons of one class. The positions and connection patterns of brain neurons are known to be set up during early life by molecules such as netrins (Moore et al., 2007) that guide neurons to specific positions and their axonal branches to connect with other neuron classes. There are several billion neurons in a brain but not billions of genes to encode unique behaviour for each cell. Using combinations of genes, you could in theory encode unique functions for every cell with just 30 genes. B lymphocytes use gene combination to start off diversity using random shuffling. I am not aware, however, of any

systematic mechanism suggested to give every brain neuron unique information. The site-specific programming of sensory and motor cortex areas, with reflexes like suckling wired in, means functional specialisation is not trivial, but it remains likely that at least thousands of neurons in a particular cortical area belong to one functional class and receive the same genetic instructions.

The gist of the above is that at each stage of computation in the brain the same information is sent to thousands of downstream neurons genetically set up to do the same sort of job. Yet it would be wasteful (in fact unworkable) if they did exactly the same job. Presumably, in development and through plasticity, each cell acquires a slightly different response profile for any particular combination of signals.

The obvious option is that even if each of a bank of a thousand cells gets inputs from the same upstream cells the detailed pattern or 'weighting' of inputs is different. A random mechanism would be a reasonable default. The immune system needs a dedicated enzyme system to switch to random antibody production because of the tight asymmetry of DNA chemistry. Growing neurons can, however, take advantage both of the spatial symmetry of surrounding tissue and of their cytosolic chemistry in generating axonal branches such that randomness of connection site is to be engineered out, rather than in. If the system needed to ensure every possible combination of signals was allotted one neuron to respond to, strictly systematic programming of connections would be needed. However, the number of possible combinations of signals is so vast that even with many millions of cells receiving the same 'broadcast' signal content there would be no chance of covering every option.

This raises an interesting aspect of our concept of freedom. In part it implies a wide range of options, ideally covering every conceivable possibility in a context, best served by systematicity. Yet we also relate it to unpredictability, which implies less than the fullest range, some options being available for some systems but not others. The living interface between fluid and solid phase provides the ideal setting for scaling up random events but the evolution of complex genetic control also provides ways to be systematic. Life makes use of both aspects elsewhere so why not in brain?

Although the detailed mechanism of individual neuronal functional specialisation is not known, we see evidence for specialisation in the work of Quian Quiroga and colleagues (2005) who found neurons in cortex that responded specifically to images of famous people or animal types. We know that neurons acquire unique response profiles, as B lymphocytes do. When presented with a scene the brain uses axonal branching to access cells committed to a vast range of possible interpretations of what is seen and the best fit can be chosen. If several cells respond it is likely that the cell with the most precise fit will fire first and can take advantage of collateral inhibition to become the salient response. If necessary, the system can re-run the process with more data or modulation of sensitivity of response until a choice emerges. This is in line with the sort of re-cycling process Doyle (2013) has suggested for a two-stage free will model.

Object recognition processes are mostly, although not always, subconscious. Consciously choosing what to do is more complicated than just recognition. Yet neurobiologically a similar mechanism is likely to apply - representing a wide range of options in term of predicted possible outcomes for competing actions and computing the best fit.

An important implication of the above discussion of neural processes is that an ultimate response is dependent on events



distributed across many neurons. This follows the neuron doctrine that all computational events in brains occur in individual cells, (however much we might sense they occur holistically). There have been attempts to challenge the neuron doctrine but the anatomy is hard to argue with. As for the B lymphocytes, we expect to have possibilities set up in separate cells and *choices* occurring as events in separate cells, or at least arising from competition between such separate events in terms of timing. Answers to questions like ‘should I choose to do H, I, J, or K’ cannot be produced by a single unit with an output that can only signal yes or no, or even a single value on a scale like rapidity of firing.

The upshot of this is that although we can build a two-stage model for free will, based on freely generating options and choosing for good reason, that belongs to a brain as a system, it is unlikely we can attribute either setting up of possibilities *or choosing* to any single ‘agent’ unit. The best a single unit can do is probably *endorse* a choice: ‘yes, I really do choose K’. This is likely to worry those who believe free will involves some sort of ‘central control agent’. A two-stage free will model probably not only implies that the will is not the free bit but that although there is a real process of choice it may not be quite what we thought of as ‘me’ that is doing the choosing!

#### Question 4: Are determinism or randomness our only options or is creativity or novelty something more?

When people such as Hume (2000) suggest that events are either determined or random, with no ‘middle way’ they may be in one sense completely wrong, but in another sense likely completely right.

Quantum theory tells us that for any given situation the probability of any particular class of event happening next is always *between* 0 and 1. Nothing is totally necessary (some things are impossible) but no event is completely random. Certain aspects of an event, such as direction of a photon or handedness of a chiral molecule, may be totally random, but the *occurrence* of the event will be governed by a probability for that context. Everything is in between random and determined.

This situation is helpful in the context of free will in that it removes the need to opt either for all choices being determined, or none, which never seemed right. We feel some choices are almost completely necessary, with others finely balanced. The common notion of free will is that with a strong reason to choose A over B, you can do so, but with no reason to prefer either you can toss a coin. Doyle talks of choices being ‘adequately’ determined and this is the key point. We can allow choices to be more random if they don’t matter.

We are increasingly familiar with a situation in which randomness seems absent – in mechanical computers. This has even come to be seen as a model of the physical world although it is actually an anomaly, with randomness airbrushed out. However, Blum and Blum (2022) show that things are more complicated, especially if you engineer back in an indistinguishable pseudo-randomness. I shall return to this.

The exclusion of a third option beyond determinacy and randomness makes more sense in terms of excluding some explanation that is neither a cause, reason, or necessity, as generally understood, nor simply a lack of those. People talk



of creativity, novelty, or inspiration as if other sorts of explanation. An important aspect of Doyle's two-stage approach is that it provides a basis for deflating a 'third way'. Van Gogh's way of painting irises might be said to be neither random nor due to any rules or reasons: just inspiration. Yet it is reasonable to suggest that, as he put paint on canvas, he had reasons for keeping some things a certain way and changing others. Reasons are close to rules and regularities and physics is just rules and regularities. Subjective reasons may appear different from the reason why chlorophyll absorbs blue photons, but can we be sure either that they are mathematically different or that they cannot be based on underlying chemical or physical reasons? In a sense the answer must be no, largely because if they are not rule based it is hard to see how to test a related hypothesis. No doubt van Gogh made repeated 'adequately determined' choices as he built up the picture, each of which produced a skilfully exact result. And if the results were not exactly determined, would they not just be a bit random?

Doyle suggests that creativity or novelty is handled well enough by the two-stage model at the first stage. Where all possibilities are open to exploitation, novelty is nothing very remarkable. A self-replicating system, such as an animal species that can throw up a vast range of possibilities from which to choose is likely also to throw up individuals that throw up possibilities never thrown up before. In the case of the immune system, it may be that autoimmune disease occurs when a possible antibody species is thrown up that subverts the choice mechanism and backfires. Several cancers may be due to the same possibility-generating mechanism (Verhest & Heimann, 2008) producing a molecular event that takes everything in the wrong direction. Ironically, Neuberger, who elucidated the antibody diversification mechanism, succumbed to such an illness.

The story might be more complicated, though. Creative minds may have some talent that is not just for throwing up more possibilities. Perhaps they have a different choosing machinery, or strategy. Amongst modern painters, both Francis Bacon and David Hockney built images most people would consider unsuitable for choosing as they went along, and many consider unsuitable once completed – maybe gruesome or trite. But they are considered first rank because there is something compelling about their choices. Hockney's hawthorn bushes may look clumsy but after seeing a few more real bushes his images can look uncannily apt. Hockney has indicated that it isn't really art unless you play with viewers' choices in this way.

Just as choosing can be real, without itself being indeterministic, because it can operate on indeterministic options, creativity is something real and valuable, but is it 'novelty'? If novelty is supposed to break the bounds of conceivable possibilities it threatens to be an oxymoron. A creative artist can only ever choose some action that is within the range of the possible. But the complexity of the range of possibilities is such that creative minds are unlikely ever to run out of possibilities to choose in ways have occurred to noone before.

This tension raises again a point made under question 3: that our concept of freedom has two competing aspects – maximum range of options and the scope for unpredictability. Creativity in a sense entails *limitation* on freedom. Complete freedom would be systematic – all options covered equally. That might be advantageous, but come a price in terms of machinery. Importantly, it may also be impossible in principle, since options for futures proliferate combinatorially beyond what any system can in practice represent. Less comprehensive coverage also provides a basis for variation in response

between individuals, making sense of a concept of creativity and perhaps also for morality? If all individuals had the same *optimal* freedom, there would seem no reason for one person to behave better than another.

A random component to a process of generating or representing options might appear to be the best way to allow unpredictable variation between individuals. However, as Blum and Blum (2022) point out in the context of computer systems strictly deterministic random number generators can generate variability indistinguishable, under real world constraints, from true randomness. Moreover, other strategies based on deterministic processes could readily be used to programme individual systems to differ in the subset of options they represent.

It seems that both premises of the traditional paradox of free will may be suspect. Rather than an entirely deterministic world being at odds with a need for free will to be indeterministic, it looks as if we live in an indeterministic world in which free will may not in principle require indeterminism. Paradoxically, it may be more the flip side of unavoidable limitations and constraints of real-world systems.

In a real world it is unlikely ever to be possible for all options for similar systems to be equally accessible, partly because of different external relations and partly because of internal structure. Freedom to be 'creative' will exist for this reason alone, but as James (1890) indicates we are particularly interested in *internal* factors. It is often said that free will is 'the ability to have done otherwise'. There are, however, metaphysical problems with that. As Leibniz pointed out, another human being, indistinguishable from me up to the point of doing, might then do otherwise, but not, considered retrospectively, me in this world, in which my identity now includes having done what I did (see Edwards, 2016). Our idea of free will may be better illustrated by two identical twins, June and Jill, where at any point June can do other than Jill, even if (hypothetically) their lives up to that point have been identical.

For something to be 'all June's fault', or, more positively, 'all credit to June', may require variation between Jill and June within, i.e., not derived from external influences. To this extent a free will with moral implications may specifically require some internally generated unpredictability. True internal randomness can achieve that by producing the 'imperfection' of option range that makes something down to June. What is confusing is that deterministic programming using pseudo-randomness can seemingly produce an indistinguishable situation in practice. Blum and Blum (2022) point out that a computer system can be built that appears to satisfy the definable dynamic requirements of free will. It needs to have a complex computational architecture and *limitation* of possibilities is an essential factor. Their system also builds on probabilistic outcomes. Yet these are achieved by what is strictly speaking deterministic pseudo-randomness.

There seems to be a principle at stake here that keeps slipping through our grasp. How could true randomness make things down to June (because it is not programmed from elsewhere) and yet pseudo-randomness programmed elsewhere be indistinguishable? Fortunately, I think we can be pretty sure that there will be true randomness in June's brain, due to fundamentals of physics, so the imponderable situation may relate only to the computer.

As Doyle (2013) indicates, the free will we consider 'worth having' may involve a more complex interplay between random possibilities and choice than historical two-stage models suggest. Variable probabilities may be involved at several stages.

Taking a step back, it may be good to remember that in modern physics the idea of any single definable A causing B has long-since been abandoned. Causation is recognised as multifactorial and distributed. It may be simply unreasonable to insist that something either is or is not 'down to June'. Which is probably why the analysis gets complicated. Nevertheless, it does seem that we can relate the concept of free will to a two-stage process that in principle could make that distinction and at least some of the time can be pinned down meaningfully.

## Question 5: What could adequately determined choices be?

Doyle's two-stage system makes our choices reason-based and 'adequately determined'. Worthwhile free will requires something like this in that to have 'freedom' one must surely be able to guarantee that an important choice can be acted on for certain. To be free is to be able to choose to remain silent under torture. It does not, however, require that all our choices are based on black and white reasons or qualitative rules. A choice of a truffle rather than a peppermint from a chocolate selection may be unmotivated, one way or the other.

In other words, a degree of indeterminacy may come in to the choosing process as well as laying out options, but in as much as it does, in inverse relation to the relevance to valued free will. A mechanism is available in that the multiple branched pathways between neurons may return equally rapid and strong responses despite all attempts to fine-tune differences. If the system is driven to make a choice, we may be in a situation similar to that created by Heisenberg's Principle, where nothing can allow us to escape allocating probabilities to more than one outcome (a solution Blum and Blum (2022) also appear to have been forced to adopt).

The first, general, conclusion one can take from this is that choosing fits well enough into the ways of the physical world with our reasons, either tightly or loosely determinate in application, behaving like fundamental physical laws, which again may be tight or loose in terms of the outcomes they predict, depending on context. A mixed solution of non-chiral molecules may 'choose' to generate right and left-handed chiral products with equal probability. A solution containing a chiral enzyme may always choose to generate a left-handed product.

The second possible conclusion is harder to judge. When a choice in a brain is evenly balanced, as for the chocolates, it seems unlikely that this is because of symmetrical probabilities for a molecular level event, even if it owes its existence indirectly to many such prior events creating multiple options. It seems more likely that macroscopic events in terms of action potentials are so closely timed or nearly equal in strength that the chance of triggering one outcome is the same as another. But that would be to miss from the analysis the critical event that finally triggers outcome. What is that going to be? We have an action potential coming in at  $t=4.97$  seconds with an amplitude of 5.26 for the truffle and an action potential coming in at  $t=4.97$  seconds with an amplitude of 5.26 for the peppermint. What gives?

The question has been phrased to imply an answer, but I think legitimately. What must give is a response to input within one or more neuronal dendritic trees that is crucially balanced to go one way or another. Introspection suggests that if there is no consistent conclusion from several passes the brain may put the question differently, to force an answer – if only because the chocolate-offerer is getting impatient. But something has to tip the narrative when the truffle is in frame

and maybe there is a default to ask if this can be implemented unless any neurons strongly object. However we think this recycling process unfolds, at some point at least one dendritic tree must host an event that flips this way rather than that.

Fortunately, we know that dendritic firing depends on a critical event in the domain of the entire tree. Much as a photon may arise from a complex field pattern, a dendritic spike can arise from a complex pattern of potentials across the entire tree. The origin of firing may be divisible into subprocesses, but it is widely recognised that there is a critical point beyond which it is not meaningfully divisible. Whether this all or nothing aspect can be treated as comparable to the all-or-nothing nature of individual quanta is moot (if long range collective field quanta are involved it might), but they may be said to share what can be called *formal* or pattern-based causation. As for a quantum such as a photon, there will be uncertainty about whether it actually occurs and exactly when it occurs, but there are rules requiring that under certain types of situation it is highly likely to occur.

The factors determining critical conditions for firing in dendritic trees are perhaps the least well worked out part of neurobiology. Firing is not just a matter of potentials adding up until meeting a threshold (Tiesinga and Sejnowski, 2009; Schmidt-Hieber et al., 2017). Various bits of evidence suggest that some patterns of signal combination are more potent than others – so-called coincidence detection (Bloss et al., 2018). A related possibility is that some form of non-linear responsiveness akin to resonance may operate. The tiniest change in a flautist's breath pressure can convert a note to an octave higher. My personal view on plausible options is given elsewhere (see Edwards, 2020; Edwards, 2023) but the question is wide open. And ultimately, this sort of behaviour will hinge on a zone of uncertainty with  $p=0.5/p=0.5$  because everything hits Heisenberg's Principle at finest grain.

In other words, we can identify critical events that might mediate choices in individual dendritic trees. What seems much more difficult to marry with known biophysics is the idea that choice events could occur at a more global system level, despite the current fashion for seeing minds in terms of complex non-linear systems.

In other words, although we do not yet know exactly how choices are made by neurons, we have every reason to think both stages of a two-stage form of free will can reflect fundamental rules of physics. Ironically, computers, although supposedly paradigmatically 'physical', are deliberately designed to appear not to do this, by only being allowed to generate 0 or 1 from 0s and 1s. The 'Conscious Turing Machine' designed by Blum and Blum (2022) is intriguing in this respect in that they found that to produce a system that generated behaviours reminiscent of a conscious organism, based on a Global Workspace architecture, it was necessary to introduce a weighted probability to outputs from modules involved in selecting what material to recycle into the Workspace. The actual outputs are determined by a pseudorandom Random Number Generator, but one that was not determinably different from true randomness. Again, the complexity of the mix of roles of indeterminacy and systematic determinacy in systems appearing to have free will is emphasised.

## A further question: Does free will require a sense of freedom?

In the context of morality and justice free will is often seen as requiring having a sense of that freedom. Criminal intent must be conscious intent. Might it be that however much the dynamics of free will are consistent with known physical

processes, without any ‘magical agent’ intervention, the crux of our concept of free will hinges on a certain type of conscious mental representation? Blum and Blum (2022) claim that their Conscious Turing Machine would have such a sense but it is unclear where within the physical dynamics this representation would arise.

At present we have no clear indication what a sense of something depends on. Although animals with a sense of something (like us) may use workspaces for thinking that does not mean such architecture confers the sense of something. It may just allow us to have a sense of a greater variety of things. A serious possibility is that very special physical events are needed to provide a sense of such things, perhaps only found in neurons. Computers may not support such events. In contrast, at the larger complex system level it is harder to see what they would lack in principle. The crucial question is at which level subjectivity occurs. We have no direct evidence on this, or even how many subjective representations a brain supports at a time.

Further exploration of this aspect is beyond the scope here perhaps we have yet to reach the smallest Russian Doll.

## General discussion: Formal cause, individual responsibility and inner teamwork

It is difficult not to conclude from the above discussion that human beings can have a form of free will or choice that satisfies reasonable requirements for the intuitive concept, without needing to include any ‘non-physical magic ingredient’ that some hanker after. What we know of basic physics and human biology can flesh out a two-stage model. We know that all causation has a random element, most evident in fluid phase. That can be captured by solid phase structures in biological systems to allow random possibilities to be subjected to tightly controlled choice events.

Choice itself sits comfortably within a modern physics treatment of causation. In modern field theory, new events arise as the most ‘fitting’ responses to complex local field patterns, their fittingness being reflected in high probability of occurrence. The detailed biology also provides a good basis for saying that free choices are the responsibility of the human being itself, in that they are not directly determined by environmental influence. Very often choices are based on hypothetical futures represented by what are probably randomly generated internal patterns.

In other words, a determinist stance, compatibilist or incompatibilist, is out of court. Libertarians have all they could hope for in indeterminate biophysics. They can insist on extra magic, but why would it be needed? Free choice can be uncontroversial reality. It arises as an instance of a ‘formal cause’ that can never exist quite like that elsewhere. If forms develop through random events – including programmed ones – formal causation arises locally. It belongs to an individual and is their responsibility. Hume (2000) was right - the man is free - but perhaps freer than he supposed.

The onus is on the extra magic devotees to say what would be extra. Biophysics offers the option to lay out a vast range of options and to select. If anything freedom and creativity seem likely to relate to limitations on options. At the fundamental level rather than determinate billiard balls we have electron orbitals ‘choosing’ to stay the same or transform into each other. For the simplest systems that will follow very predictable rules but as soon as we have complex biochemistry all sorts of competing ‘preferences’ will come into play. Discussion of free will sometimes seems to highlight

the option to be perverse or illogical – to choose to read the bible underwater maybe. But once we get to the level of chemical complexity of a nervous system there seems no particular reason why the motivation should not get complex and conflicting. Such conflicts are now easy enough to programme into artificial intelligence.

Except that there are less comfortable questions remaining. Who exactly does choice belong to? The whole organism with its narrative is entirely responsible because the choice arises from a uniquely posed range of possibilities. The whole organism is, however, a complex network of separate events in spacetime within which there is little chance of identifying any central ‘controlling agent’. Nor is there good reason to identify the site of choice with a unique experiencing subject. In a Global Workspace architecture (Blum and Blum, 2022) we should expect the computational units receiving ‘broadcast’ content to be the sites where being conscious of freedom occurs – and there will be a lot of them.

The initial problem is that the ‘free’ nature of the process is in large part down to events involved in setting up representations of multiple possibilities that are likely to be distributed widely with the nervous system and very probably in structures not directly involved in either conscious experience or choice events. We could say that we accept that the ‘chooser’ is provided with resources in this way but that we still want to identify this chooser, as the owner of the ‘free will’. In physics terms the answer is probably austere, reminiscent of Locke’s deflationary account of an apparently enduring self as simply events linked to others through memory. A fundamental causal event in modern physics is a field quantum arising (then annihilating) in the context of a local field pattern. Other quanta may come and go in the same event as in a so-called Feynman diagram (see Feynman et al., 1964). The ‘agent’ in this event is in a sense *the local field pattern that gives rise to it*. One could say the local pattern chooses that event to occur. Whether or not events in neural dendrites likely to be involved in ‘choice’ responses by generating spikes fall under a quantum-theoretic analysis is uncertain, but, as indicated above, the general form of dynamics is likely to be at least analogous. How many choosing domains there are in a brain is anyone’s guess but very likely a lot. However much current neuroscience may make a pig’s ear of explaining conscious thought, the rejection of Descartes’s single experiencing and choosing soul is based on good evidence.

Isolating choosers also has downsides. A chooser that did not generate possibilities to choose from may not be fully responsible for what it chooses. It may also in a sense not be responsible for the way it chooses if that part of the story follows general rules of fundamental physics, as it might well do. Free will worth having seems to be a matter of teamwork.

This again highlights the issue of whether or not random development of neuronal responsiveness in June’s brain is down to June as a person or agent. Some might suggest that June as an agent is something separate from her material biology – which might make the randomness issue irrelevant. But neither of the two apparent options seem to work. If June as a person is considered as a ‘whole system’ or ‘organism’ then she is responsible for everything going on in her brain, including random neural responsiveness. If she is some additional agent that overrides the brain events then there is no known available mechanism and the claim of free will just hangs on an introspective intuition.

In short, although we can dispel any suggestion that science needs to deny the existence of meaningful free will, some libertarians are likely to find some of the associated implications unpalatable!

The consistency of free will with the most fundamental aspects of all physics has been emphasised. However, that is not to deny that the free will we like to think we have is special. There is good reason to think our sort of free will is a remarkable dynamic property likely unique to organisms with complex nervous systems (to represent possible *futures* to choose from). Similar dynamics may now be achievable in computers but there remains a question about having a conscious sense of freedom.

This brings us back to the central structural feature of Doyle's two-stage model. The division into two-stages breaks the impasse in providing a coherent account of 'free will'. It may seem to separate freedom from will. However, that freedom arises essentially from the combination of discrete events and a continuous symmetric dynamic metric (with infinitesimal precision disallowed), which applies in biological systems at all levels. In computers it might normally be engineered out but in practice for complex systems something very similar can be put back in (Blum and Blum, 2022).

In conclusion, there is no case for saying that free will as we can describe it is outside the scope of physics. Rather than physics being boring Laplacian billiard balls, it appears to provide all the ingredients freedom and willing require. If it is claimed that 'freedom' or 'creativity' are by definition outside any rule system physics might use then the onus is on the claimant to explain what the extra magic would provide and why physics cannot do it.

## References

- Bloss, E.B., Cembrowski, M.S., Karsh, B., Colonell, J., Fetter, R.D., Spruston, N. (2018). Single excitatory axons form clustered synapses onto CA1 pyramidal cell dendrites. *Nat. Neurosci.* 21(3):353-363.
- Blum, L. and Blum, M. (2022) A theory of consciousness from a theoretical computer science perspective: Insights from the Conscious Turing Machine. *PNAS* 119(21) e2115934119.
- Born, M. (1926, 1983). "I.2". In Wheeler, J.A., Zurek, W.H. (eds.). *Zur Quantenmechanik der Stoßvorgänge. Zeitschrift für Physik.* v37, Princeton University Press. pp. 863–867.
- Dennett, D. (2003). *Freedom Evolves*, London, Viking Books.
- Doyle, R.O. (2013). The Two-Stage Solution to the Problem of Free Will. In Antoine Suarez, Peter Adams (eds.), *Is Science Compatible with Free Will?*, New York: Springer. pp. 235-254.
- Eccles, J.C. & Popper, K.R. (1984). *The Self and its Brain*, Abingdon, Routledge
- Edelman, G.M. (1987). *Neural Darwinism: Theory of Neuronal group Selection*. New York, Basic Books.
- Edwards, J.C. (2020). Quantum-level experience in neural dendrites: an interpretation-neutral model. *J. Consciousness Studies*, 27 (11-12):8-29.
- Edwards, J.C. (2016) Leibniz and Telicity. In *Proceedings of the 10th International Leibniz Conference*, vol 1. Hannover, University of Hannover.
- Edwards, J.C. (2020). Quantum-level experience in neural dendrites: an interpretation-neutral model. *J. Consciousness Studies* 27, 11-12, 8-29.
- Edwards, J.C. (2023). Modern Monads, Leibniz, Continuity, and the Stream of Consciousness *Qeios*, W5X50X.
- Feynman, R., Leyton, R. and Sands, M. (1964, 2005). *The Feynman Lectures in Physics*, Boston: Addison-Wesley.



- Heisenberg, M (2009). Is free will an illusion? *Nature*, 459:164.
- Hume, D. (2000). *A Treatise of Human Nature* (Norton, D.F. and Norton, M.J. eds.), Oxford/New York: Oxford University Press.
- James, W. (1890/1983). *The Principles of Psychology*, Cambridge, MA: Harvard University Press.
- Moore, S.W., Tessier-Lavigne, M., Kennedy, T.E. (2007). Netrins and their receptors. *Adv. Exp. Med. Biol.* 621:17-31.
- Neuberger, M.S. (2008). Antibody diversification by somatic mutation: from Burnet onwards. *Immunol. Cell Biol.* 86(2):124-32.
- Poincaré, H. (1905, 1958). *The Value of Science* XI, New York, Dover Publications.
- Quian Quiroga, R., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature* 435:1102–1107.
- Schmidt-Hieber, C., Toleikyte, G., Aitchison, L., Roth, A., Clark, B.A., Branco, T. and Häusser, M. (2017). Active dendritic integration as a mechanism for robust and precise grid cell firing. *Nature Neuroscience* 20(8):1114-1121.
- Schrodinger, E. (1944, 1992). *What Is Life?* Cambridge, C.U.P.
- Tiesinga, P. and Sejnowski, T.J. (2009). Cortical enlightenment: are attentional gamma oscillations driven by ING or PING? *Neuron* 63(6):727-32.
- Verhest, A. and Heimann, P. (2008). Basic Cytogenetics and the Role of Genetics in Cancer Development. In *Comprehensive Cytopathology* (Bibbo, M., Wilbur, D.C. eds.) Philadelphia, Saunders.