A Model for Free Will

In this model, I attempt to provide a framework for free will that has the following characteristics:

1) There is a sense of "true choice" that goes beyond simple lack of predictability.
2) There is a clear *entity* that makes the choice.
3) There is a sense of *responsibility* or *agency* to the choice.

Within this context, I am explicitly not considering either decision making by randomness, such as simple quantum indeterminacy in the decision process, or any ultimately algorithmic element of the choice to be free will.

The motivation for this attempt at modeling is the deep, intuitive sense that I have that there are moments where I am making a choice that goes beyond the actions of my physical constituents obeying the laws of physics (or laws of nature). But, when I do make such a choice, my physical constituents certainly do not violate the laws of physics either. Additionally, there is a recognition that many, if not most, of my actions are the result of my physical constituents obeying the laws of physics, and in fact, any true choice must be compatible with such a description of my constituents.

Though not universally accepted as the way the world actually is, as a first step, one can fairly easily imagine an entity (consciousness, etc.) that is the agent of choice and the choices it makes are always compatible with the laws of physics. The simplest way to understand this is through the concept of *emergence*, where there is a higher level description that encompasses the concept of free will. However, even when accepting the possibility of emergence as a starting point for understanding free will, there is a central question that must be answered if one is to develop a model of free will that fully captures choice in its deepest sense.

Can you have a cause without algorithmic laws? For there to be an impact of my choice, it must be able to cause changes to the physical constituents that are consistent with the laws of physics. This suggests the existence of an additional "law of nature" at work that will be algorithmic and contradict my attempt to have a free will that allows for true choices that are not constrained by an algorithmic law (even a statistical/probabilistic one). This challenge requires a more careful consideration of three concepts:

1) Concept of "cause"
2) Concept of "emergence"
3) Concept of time

Concepts of Cause

The most intuitive understanding of a cause is captured by Newton's Second Law: Forces cause a change in momentum (or for constant mass, forces cause an acceleration). Most of our current understanding of physical systems has this general conception of causes of motion at its root because of how well it maps to any experimental measurements of physical systems. Of course, when one focuses on the more "fundamental" level of quantum phenomena, the description of

"forces" changes. However, our modelling of a system still focuses on the *interactions* between physical objects, and we characterize some objects as "force carriers". For example, the photon is the particle that mediates electromagnetism. Therefore, there is an entire set of physical laws that exhibit this intuitive sense of cause, and embodies physical objects interacting with each other and changes occurring due to those interactions.

There is another type of physical law that is best exhibited by statistical mechanics. In this case, one is considering systems with a large number of degrees of freedom. A great example is the Second Law of Thermodynamics related to the increase in entropy of a system. This is a constraint on a system's behavior, and it is always in the context of the individual particles still obeying the fundamental laws of interactions. But, the Second Law of Thermodynamics is a stand-alone principle that is not derivable from the underlying laws.

In the context of thermodynamics, something that is directly related to our attempt to understand causes is the concept of a phase transition and the free energy. In this case, the explanation of the thermodynamic state of the system is the minimization of the appropriate free energy, which is basically a balance of energy and entropy considerations, with a focus on *global* variables for the system, such as temperature, pressure and volume. In this framework, one causes a change in the state (e.g. fluid to solid) by changing a global variable (e.g. temperature).

From the perspective of the physical objects, the cause associated with a phase change does have a different character than the cause associated with the interaction between physical objects. The key concept is the *system and its boundary*. Though one can think in terms of external and internal forces when considering the nature of interactions, the *cause* of interest is the *direct* interaction with the particle of interest. Also, there is a fundamental symmetry in the interaction. For example, however we picture the interaction between two particles through forces, Newton's Third Law requires an equal and opposite interaction. In a quantum force particle model, the two objects that interact *both* exchange virtual photons, and so on. For our thermodynamic model, the cause can be asymmetrical. For example, one can have a sealed container of fluid with a conducting wall that is heated uniformly (perhaps with electrical heaters in the wall). From the perspective of the phase change caused by this temperature change, there is not necessary a symmetric impact on the external system heating the fluid.

The concept of the *boundary* as the cause of change is even more apparent in the behavior of complex fluids. Complex fluids are systems of multiphase components (sand particles in air, gas bubbles with liquid walls, i.e. foam, liquid bubbles of one type with liquid walls a different composition, i.e. emulsions, etc.). A key feature is that the "mechanical state" (solid that holds it shape versus fluid that flows) is not determined by the thermodynamic state of its constituent molecules. Instead, it is determined by boundary conditions. For example, in a foam, the molecules are all free to flow (gas bubbles, liquid walls), but the collection of bubbles that make up the foam hold their shape and are considered to be in a solid state. If you apply enough shear force or strain the system long enough (deforming the *boundary*) the system will make a transition to a fluid state and begin to flow. This is a non-equilibrium state transition completely caused by a change of the boundary. Once again, the individual bubbles and molecules are not violating any laws of physics, but their ultimate behavior is determined globally at the boundary of the system.

In the case of the foam, it is very clear that the foam and the external system driving the boundary are fundamentally different systems. For many cases of particle interactions, one gains different insights into including the interacting objects as a single system or as two different systems. For example, a particle interacting with the Earth when seen as its own system will be described as "having the force of gravity due to the Earth" acting on it. If we take the Earth and the particle as a single system, we discuss the exchange of "gravitational potential energy for kinetic energy". These are equivalent pictures of the same behavior that provide additional insights into the nature of interactions, and provide different computational power depending on the situation. For the foam being driven by a shear, there is no change in the description with the inclusion of the driving mechanism as "part of the system" or "external to the system" other than semantics. The same formalism and modeling is applied to the foam – it is just either called "the system" or the "subsystem"!

Emergence

In our discussion of causes, we encountered two excellent examples of emergence: entropy and foams! The basic idea behind emergence is that systems with sufficient level of complexity exhibit new behavior or laws that are not derivable from the underlying fundamental particle interactions. One might qualify this to be a "strong form" of emergence, because I specified that the new laws are "new" and not derivable from the lower level laws. There is also a sense in which one has higher level descriptions that are derivable from the lower level laws, but are a more efficient description of the complex phenomena.

To clarify the derivability of emergent phenomena, one might consider the difference between the Second Law of Thermodynamics and the Navier-Stokes equation. The Navier-Stokes equation provides a coarse-grained description of fluid flow based on "fluid elements" and is effectively Newton's Second Law for fluid elements. Therefore, one might argue that it is not a "new" law in a fundamental sense. In contrast, the Second Law of Thermodynamics (as noted by Landau and Lifschitz in their Statistical Mechanics book) is a law of a different kind. It fundamentally depends on the existence of many degrees of freedom and the particular types of ensembles and measures of phase space that are relevant to physical systems. In other words, statistical mechanics is not derivable from two-particle interactions, though it is always consistent with those laws.

Concept of Time

For the purposes of this discussion, the relevant issue of time I would like to point to is the growing evidence from physics that we need to distinguish between "reality that just always has been" and "our experience of the evolution of space-time". First, experiments to test general relativity strongly support the idea that space-time is a physical thing that evolves, and our particular space-time in which we live had a specific beginning. Coupled with this is the concept of a reality that just has always existed and is "outside of time". There are a number of different potential aspects for this reality, but the two most relevant features for our discussion are:

1) The concept of existence without time.
2) The separation into physical and non-physical aspects of reality

In this context, the part of reality described by space-time (and the particles and fields interacting within space-time) is what I will refer to as physical reality (and there may be more than realizations of space-time). Other concepts, such as a quantum vacuum state from which space-time emerges as a fluctuation or an initial four dimensional space that had a transition in one direction to a time dimension are examples of constructs that I would argue are effectively non-physical as they are not accessible to measurements and study in a direct physical way.

Building the Model of Free Will

Starting from the concepts of cause due to boundaries and emergence, we can build the following model for free will.

1) A physical system that is sufficiently and appropriately complex will have as an emergent property some form of consciousness in the sense that it can be identified as a coherent object that is an "I".
2) The physical elements of this particular system will have a well-defined boundary.
3) The key element is that this boundary is a boundary between the *physical system* and some aspect of *non-physical* elements of reality.
4) The "I" making the choice is ultimately associated with both the physical systems, its boundary, and some part of the non-physical.

What is gained by this view of free will? First, it suggests a mechanism for a cause that is not necessarily constrained by algorithmic laws in the following sense:

1) A choice occurs when there is a change in the boundary condition between the physical system identified as the "I" making the choice that causes a state change in the physical system.
2) The change in the boundary condition of the physical system produces a change of state in the physical system that is consistent with the laws of physics.
3) The non-physical region that interacts with the physical through the boundary can ONLY interact with physical systems by changes in such boundaries.
4) Therefore, one can ascribe a concept of "will" to the non-physical that can "cause" changes in boundaries that create changes in the state of a physical system, and "will" can be an inherent, non-algorithmic property of the non-physical system that is well defined but not subject to algorithmic laws. It just exists. In the comments below, I refer to this property as "agency". Not sure what is the best word for this.

This model also allows for "degrees" of free will within physical systems and for free will to evolve in systems due to natural process.

1) The "degree" of free will could be connected with the "nature" and "amount" of boundary with the non-physical
2) The emergence of a boundary will evolve as complex systems evolve and develop new interactions and behaviors.

There is an interesting analogy here to the evolution of behavior in computer simulations. The physical hardware is clearly the collection of electrons in the silicon chips interacting according to the laws of physics. The information generated in a simulation emerges from the interaction

between the rules laid out in the programming code and the rules of physics governing the physical hardware. This is an example of changes in states of a system due to interaction across a boundary of two fundamentally different realities with different sets of rules.

Notice, the critical role of a boundary between two *fundamentally different systems* which is the analogy to the foam, its boundary, and the system that drives it. Even though the foam analogy is about two physical systems, the point is that they have very different properties. In this model of free will, we need a system with "agency" as a fundamental property in a way that does not seem possible in physical systems.

Implications

One of the core issues I believe this model address is the need for "agency" as a characteristic of reality if one is going to truly have free will. In this case, I am putting "agency" in a similar category to something like charge. At one level, charge does not need to be explained. It is that property of matter that mitigates electrical forces. It just exists. Agency could have that character if there was not the challenge of the fact that it does not appear to follow a "natural law" and results in "arbitrary" choices. (I am using arbitrary in the hopes of distinguishing from unpredictable in the sense of probabilistic or statistical, but with an underlying natural law governing the statistics.) Given how the fundamental particles/interactions of physics reality do not appear to have any agency, it is difficult to imagine how it can be an intrinsic feature of physical things.

One could imagine a model where agency is a property of the non-physical aspects of reality and can cause any possible action. Intuitively, this seems problematic because it drops the concept of any rules and is purely a "magical" solution. The key element of the proposed model is the role of the boundary in defining and limiting the vagueness of "agency" as a property of reality, even if it is a property of non-physical reality, by providing a clear, and well-defined mechanism for impacting physical reality. There is still the problem of what it might mean for non-physical reality to have a property called "agency", but we have reduced the scope of the problem significantly and it is closer to the problem of why does reality exist at all?

This model also deals with what I see are challenges associated with emergence in a purely physical based reality as a means of achieving true choice (agency) that includes both an effective cause of change and the absence of an algorithmic rule that logically prevents true choice. In any discussion of emergence in a purely physical reality as the source of free will that I have read, there really seems to be a leap from "emergence seems to explain a range of interesting things" so emergence must explain free will. What is missing is how to truly handle the absence of algorithmic rules that ultimately determine the choice. It seems to me that the existence of a boundary is the first essential piece as it provides a clear definition for the "type of cause" that is most consistent with choice. Once you have that mechanism, there may be a way to consider "agency" within purely physical systems emerging. However, the simplest model seems to be the one proposed:

1) Agency is an intrinsic property of non-physical reality.
2) Agency is non-algorithmic choice, but the interaction between agency and physical systems is not arbitrary. It only occurs at boundaries with specific properties.

3) The necessary boundaries for agency are an emergent property of physical reality and occur as part of the natural evolution of systems. In principle, such systems could be designed by other systems with sufficient understanding of the state required to generate the emergent properties.
4) This explains how one has agency in certain physical systems, degrees of agency, and agency that is compatible with physical law and appropriately constrained.

A final comment. It seems that one can have “meaning” and “responsibility” as emergent properties of physical systems without necessarily having the situation proposed here. In this regard, one could have “apparent” choice, with a well-defined “I” that is consistent with responsibility even if all decisions ultimately came from algorithmic processes and there was not true choice in the sense defined here. In other words, there is still the experimental question of have to distinguish true choice from other situations of decision making given that by definition decisions are one time events that are not really subject to experimentation from this perspective.