# Dignity, Free Will, Emergence, and Illusion

*Bartlomiej Andrzej Lenart*

**Abstract**: This paper argues that although emergentism does not appear to provide the libertarian with the tools she requires for metaphysical freedom, this does not actually matter for the grounding of a robust notion of moral responsibility. Moreover, illusionism about metaphysical freedom offers some consolation to those who see metaphysical freedom as a source of human dignity and value. This paper argues that emergentism, even in its weak form, when buttressed by both philosophical and psychological considerations regarding the illusory nature of the phenomenal experience of metaphysical freedom, as well as the Strawsonian notion of reactive attitudes, serves as a source not only of value and dignity, but also of a robustly grounded conception of moral responsibility.

**Keywords**: emergence; illusionism; free will; libertarianism; determinism; metaphysical freedom; moral responsibility; dignity; human value.

## 1. Introduction

While philosophers will certainly not cease their ontological explorations and musings, various forms of materialism (I will refer to the set of these different formulations generally and broadly as *physicalism*) have taken root both in the physical sciences and in the standard approaches to many metaphysical problems. Physicalist worldviews can exude poetic imagery about our materialistic origins, as evidenced by the following quote:

> Life begins with the process of star formation. We are made of stardust. Every atom of every element in your body, except for hydrogen, has been manufactured inside stars, scattered across the Universe in great stellar explosions, and recycled to become part of you. (Gribbin, J., Gribbin, M., 2001, p. ix)

Physicalism's tendency towards mechanistic worldviews, however, also raises questions regarding our agency, dignity, and metaphysical freedom, as Drew McDermott's ominous question suggests:

> As far as science is concerned, people are just a strange kind of animal that arrived fairly late on the scene. When you look at the details of how they work, you discover that, like other life forms, people's bodies are little chemical machines. Enzymes slide over DNA molecules, proteins are produced, various chemical reactions are catalyzed. Molecules on the surfaces of membranes react to substances they come into contact with by fitting to them and changing shape, which causes chemical signals to alter the usual flow of events, so that the machine's behavior can change as circumstances change. Traditionally there was one big gap in this picture: the human mind. The mind was supposed to be a nonphysical entity, exempt from the laws that govern the stars, the earth, and the molecules that compose us. What if this gap closes? What if it turns out that we're machines all the way through? (McDermott, D. V., 2001, p. 1)

There is more at stake than just the prospect of existential angst if such descriptions of our natures are accurate. Physicalism, though appealing and perhaps even beautiful in some ways, does not appear to be compatible with free will (by 'free will,' I mean something closer to a libertarian conception of freedom than a mere compatibilist approach; I will follow Peter van Inwagen (1998) in referring to this notion of freedom as 'metaphysical freedom'). As I will argue, if we assume physicalism, then neither a deterministic physical universe nor an indeterministic one is truly conducive to metaphysical freedom without the postulation of some radical causal mechanism. See Lenart (2022) for an argument outlining metaphysical compatibilism, which proposes a compatibilist position that is logically compatible with van Inwagen's notion of metaphysical freedom.

Why should we even care about metaphysical freedom? There are many answers to this question, but one which resonates equally with most philosophers and lay people is the one provided by Robert Nozick (1981), in his *Philosophical explanations*, where he states that "[d]eterminism seems to undercut human dignity, it seems to undermine our value" (Nozick 1981, 291). Aside from the concern with our dignity and value, there are also worries with our notion of moral responsibility and accountability. Thomas Nagel writes:

> If one cannot be responsible for consequences of one's acts due to factors beyond one's control, or for antecedents of one's acts that are properties of temperament not subject to one's will, or for the circumstances that pose one's moral choices, then how can one be responsible even for the stripped-down acts of the will itself, if *they* are the product of antecedent circumstances outside of the will's control. (Nagel, 1976/1979, p. 35)

Physicalism, however, is not limited to the reductive mechanistic dogma suggested by McDermott (2001). Emergentism, which, in its classical form, is a physicalist view, provides a *prima facie* reason to lay our worries about dignity, value, and moral responsibility to rest in virtue of the fact that strong emergentist theories appeal to downward causation as a source of the autonomy of macro-level phenomena (like minds). As this paper will

University of Calgary, Libraries and Cultural Resources
Calgary (Alberta), Canada
email: bartlomiej.lenart@ucalgary.ca

Philosophical Readings XV.1 (2023), pp. 1-12.
DOI: 10.5281/zenodo.8283266

discuss, however, the concept of downward causation is riddled with serious problems.

This paper argues that although emergentism does not appear to provide the libertarian with the tools she requires for metaphysical freedom, this does not actually matter for the grounding of a robust notion of moral responsibility. Moreover, illusionism about metaphysical freedom offers some consolation to those who see metaphysical freedom as a source of human dignity and value. This paper further argues that emergentism, even in its weak form, when buttressed by both philosophical and psychological considerations regarding the illusory nature of the phenomenal experience of metaphysical freedom, as well as the Strawsonian notion of reactive attitudes, serves as a source not only of human value and dignity, but also of a robustly grounded conception of moral responsibility.

## 2. The Problem of Metaphysical Freedom

Owen Flanagan explains that "[o]ne reason…that the belief in a nonphysical mind took hold is because thinking of the mind as nonphysical fits well with thinking of human agents as free. Physical things obey natural laws, non-physical things don't" (Flanagan, 2002, p. 102). However, if Gribbin & Gribbin (2001) are correct in endorsing physicalism on the grounds that human beings are made of the heavy elements produced in stars, then the notion of a non-physical mind becomes quite problematic (this is in addition to the problems Cartesian dualism faces). If "[s]tardust is the key to the existence of complex molecules in the Universe, and therefore to the existence of life itself" (Gribbin & Gribbin, 2001, p. 181), then we are purely physical beings and, although we are very complex physical systems, we too must be governed (just as the stars, the earth, and the molecules that compose us are governed) by the natural laws, which determine the motions of other simpler (and, in fact, all) physical systems.

Flanagan, like all his fellow compatibilists, attempts to reconcile physical determinism with freedom of the will. He writes:

> [T]he basic idea is this: accept that deliberation and will exist and that they are often proximate causes of behavior. Concede, however, that they themselves are natural phenomena, indeed that they are brain processes—subject to whatever causal laws govern proximate brain causation. So deliberation and volition exist. (Flanagan, 2002, p. 114)

What Flanagan is hinting at, however, is not metaphysical freedom, but rather *self-control* defined in the following manner: "[f]or some integrated system *S*, some *subsystem Sa controls subsystems* $S_1$ *... Sn* if the relation between *Sa* and $S_1$ *... Sn* is such that *Sa* can *drive* $S_1$ *... Sn* into states that *Sa wants* them to be in" (Flanagan, 2002, pp. 115-116). The difference between self-control (as defined above) and metaphysical freedom, as I understand it and wish to define it, is that whereas the type of self-control defined above is ultimately dependent on and determined by the laws of nature (and thus the controlling subsystem *Sa*, although in control of subsystems $S_1 \ldots Sn$, is determined by conditions beyond its control just as surely as the controlled subsystems $S_1 \ldots Sn$ are determined by *Sa*), metaphysical freedom is the freedom to do otherwise even given the same initial starting conditions. Whether metaphysical freedom is possible is a separate question (one I wish to pursue in light of emergentism). However, if human beings are to be attributed with free will, it must be metaphysical freedom and not a compatibilist redefinition of the concept of free will that we have in mind when we state that some agent *S* was *free* to have done otherwise.

What precisely is wrong with compatibilism and how exactly does it differ from the notion of metaphysical freedom defined above? Some compatibilists maintain that the ability to do otherwise can still be accounted for in a deterministic world. "According to the advocates of this argument—let us call them 'conditionalists'—what statements of the form…S could have done X *mean is*:…If S had chosen to do X, S would have done X" (van Inwagen, 1975/2001, p. 27).

I think that the problem with conditionalist notions of 'could have done otherwise' is the following: assuming the thesis of determinism, following van Inwagen's definition of 'determinism,' (according to van Inwagen (1975/2001), the truth of determinism is contingent on the laws of physics: they must be precise and not probability-driven or statistical) if the conjunction of a certain state of the world *A* with the laws of physics *L* entails the state of the world *B*, then given *A* and *L*, *B* necessarily follows. If the above is true, then it cannot be the case that one possible world $P_1$ (where the laws of physics consist of the set of laws *L*), is in state *A* at time $t_1$, and state *B* at $t_2$, while another possible world $P_2$ (which is identical to $P_1$ and where the laws of physics also consist of the set of laws *L*), is in state *A* at time $t_1$, and state *D* at $t_2$. Therefore, saying that 'if *S* had chosen to do *X*, *S* would have done *X*' amounts to saying that although *A* occurs at $t_1$ and *B* comes about at $t_2$ in $P_1$, if *C* had occurred at $t_1$, then *D* would have come about at $t_2$ in $P_2$. It is like saying that if we run a system governed by a set of laws *L*, starting it in an initial state *A* at $t_1$, it will go into state *B* at $t_2$, but if we reset the system and then run it again, starting it in an initial state *C* at $t_1$, it will go into state *D* at $t_2$. Although the above is true, I do not see how it captures the sense of 'could have done otherwise,' which is necessary for genuine alternative possibilities to be open to an agent.

If determinism is truly not compatible with metaphysical freedom (and I think that the two are *not* compatible in the way most compatibilists propose), can indeterminism do the job? The difficulty with indeterministic accounts of free will is that such views encounter the problem of arbitrariness. That is, undetermined events appear to be too "chancy" to amount to the kind of control necessary for the attribution of free will to a system. Peter van Inwagen's concern with the chancy-ness of indeterministic causation echoes this worry. He writes:

> If the laws are indeterministic, then more than one future is indeed consistent with those laws and the actual past and present – but how can anyone have any choice about which of these futures becomes actual? Isn't it just a matter of chance which becomes

actual? If God were to "rerun" an indeterministic world to precisely its states at some time in the past, and then let the world go forward again, things might indeed happen differently the "second" time. But then, if the world is indeterministic, isn't it just a matter of chance how things *did* happen in the one, actual course of events? And if what we do is just a matter of chance – well, who would want to call that freedom? (van Inwagen, 1998, p. 370)

The problem of metaphysical freedom in light of the adoption of physicalism appears to be unsolvable. It may very well be the case that we lack metaphysical freedom and that the only type of control we have is the kind described by compatibilists (which ultimately amounts to the conditionalist proposal mentioned above). Assuming, then, that the universe is deterministic, what precisely contributes to the problem and what is needed for metaphysical freedom to become plausible? I am assuming determinism for several reasons: (1) Many emergentist views assume determinism to be true, (2) determinism provides us with a valuable contrast with metaphysical freedom and thus drives our intuitions about how freedom of the will should be defined, and (3) indeterminism poses similar problems for the formulation of a satisfying account of free will, but it complicates matters slightly more than is necessary for my purposes in this paper.

Physicalism is often associated with reductionism. That is, all physical objects and systems must ultimately reduce to their physical constituents or parts, which ultimately determine the properties and activities of the macro-level objects or systems. However, physicalism allows for a much wider spectrum of possibilities than just the set of reductionist materialist views, which is merely a subset of a broader collection of physicalist ontologies. Emergentism is also a physicalist proposal, but it does not put forward a reductive understanding of the world and thus, may be conducive to the libertarian notion of metaphysical freedom.

The concept of emergence, very briefly summarized, states that an emergent property (1) must be novel (qualitatively different from its constitutive parts), (2) unpredictable, (3) cannot be instantiated at any lower level (cannot exist at the level of the parts that constitute it), (4) must result from an interaction between its lower-level constituents, (5) has to be holistic (that is, an emergent property must be a property of the entire system and not just a property of its parts), and (6) emergent phenomena should, at least on the classical accounts of emergence, have the capacity to exert a causal influence on their lower-level constituents. The notion of causal efficacy of emergent phenomena appears to be quite promising for libertarian accounts of freedom. Causal efficacy of higher-level properties and their influence on lower-level parts requires that higher-level properties have direct downward causal influence on the lower-level properties of the parts that give rise to the higher-level phenomena. Although the notion of macro-determinism appears to be somewhat paradoxical, downward causation is the only feature of emergent phenomena capable of providing a satisfying account of metaphysical freedom in a purely physical universe.

The reason why unpredictability cannot offer a satisfying account of freedom, whereas the notion of downward causation can, is that while the notion of downward causation points to the possibility of actual and genuine causal influence, unpredictability is merely an epistemic constraint. To better understand this claim, let us consider Karl Popper's argument that complete prediction in a classical, deterministic system is impossible if the predictor is part of the system. Popper (1950) considers a mechanical system *A* and a predictor *B* attempting to predict *A*. *B* can only predict *A* if it can calculate its interference with *A* (this is because *B* is a part of the system it is trying to predict). In other words, *B* must include its act of predicting *A*, and the consequent effect on *A* of such predicting, in its prediction of *A*. One way for *B* to compute its interference with *A* is to study its interfering parts *B′* and their interaction with *A*, but this implies that *B* now needs to study the system *A* + *B′* instead of *A* and the same problem arises again. However, although *A* cannot be predicted by even an omniscient super-calculator if the super-calculator happens to find itself within the system (that is, within the world that contains *A*), from the point of view of a timeless, space-less, and omniscient being (a being in existence outside of the system), the behaviour of *A* would be perfectly predictable; in fact, the behaviour of *B* attempting to predict *A* would likewise be perfectly predictable (it does not matter, for the purpose of the thought experiment, that such a being is implausible). Thus, the libertarian vision of freedom requires that emergent phenomena exert a direct causal control on their underlying parts and not merely that emergent phenomena be unpredictable since metaphysical freedom demands control (which downward causation is capable of providing) and does not really (or at least not merely) require unpredictability (especially if unpredictability occurs in a deterministic universe).

## 3. Downward Causation, Strong Emergence, and Weak Emergence

The mechanistic dogma states that everything is just a complex of physico-chemical properties and laws, which account for all higher-level properties and laws (and where such higher-level properties and laws are reducible to the lower-level physico-chemical parts that constitute them). The mechanistic dogma implies that the mind is fully reducible to its physico-chemical composition and thus is fully determined by its micro-constituents. Tim Crane writes: "[t]he central idea of the mechanical view of the mind is that the mind is a part of nature, something which has a regular, law-governed causal structure" (Crane, 2003, p. 130). The mind (a higher-level phenomenon), which is arguably necessary for metaphysical freedom, on such a view, is incapable of the type of control required by libertarian theories of free will.

Supernaturalism states that life and the mind are governed by, and thus can only be understood in terms of, supernatural forces (e.g. an *élan vital*, entelechies, etc.). Supernaturalism, however, does not appear to be compatible with physicalism and thus, given the empirical and epistemic strength of our well-established scientific theories, it lacks coherence with our already established understanding of the world, which makes supernaturalism ultimately unconvincing.

Emergentism rejects both the mechanistic dogma and supernaturalism. Emergentists do not postulate super-

natural entities, but they also reject the reductive agenda of the mechanists. Because emergentism accepts physicalism without the mechanist's reductionism, emergentists can (and often do) argue that higher-level phenomena are autonomous and that they are capable of influencing their lower-level parts. Emergentism, then, appears to be just what the libertarian needs and emergence is able to provide it without resorting to supernaturalism.

If emergence is to be the key to a libertarian conception of free will, then downward causation should be a feature of an emergentist account. Without downward causation, macro-level emergent properties are dependent on, and influenced by, their microstructure and not the other way around. In either case (no matter whether determinism or indeterminism proves true), if physical bodies are governed exclusively by a micro-dynamic, then any complex organism's actions are dependent on the micro-level interactions of the organism's micro-parts.

The British Emergentists argued for several features characteristic of emergent phenomena: one salient feature is the causal efficacy of the macro-level emergent phenomena and another is the unpredictability of such phenomena. According to Mark A. Bedau (1997), there are two hallmarks of emergent phenomena, namely that (1) emergent phenomena are constituted by and generated from underlying processes (that is, the emergent macro-states of a system are structural properties constituted entirely out of the system's microstates) and that (2) emergent phenomena are autonomous from underlying processes. Bedau, in "Weak emergence," calls these types of views (accounts of emergence that postulate downward causal influence of macro-level on micro-level properties) *strong* emergent views and distinguishes them from his own *weak* emergent proposal.

What differentiates weak emergence from strong emergence, then, is the weak emergentist rejection of downward causation and the views about unpredictability. With regards to the latter, Bedau states that weakly emergent phenomena are, in principle, predictable, but that they are derivable only via simulation.

According to Bedau, derivations that depend on simulations are:

> (1) massively contingent (they include a great deal of accidental information about the system's components and the external conditions).
> (2) For this reason, derivations can be too detailed and unstructured for anyone to understand how they work.
> (3) Finally, such derivations can obscure simpler (macro-level) explanations of the same macrostates, which occur in systems with different microdynamics and different external conditions.

Bedau states that the derivation of weakly emergent macrostates requires simulation even if the predictor is a powerful super-calculator (e.g. a Laplacian demon). A super-calculator would have a decisive advantage over human beings in calculation speed, but the derivation would still be based on simulation.

The characteristic mark of *strong* emergent phenomena can be captured by means of four conditions, which need to be satisfied if a property is to be termed an "emergent property." (1) An emergent property *P* is, in principle, unpredictable from its lower-level microstructure even if one is given a complete theoretical knowledge of the microstructure, including the initial conditions and the micro-dynamic, such as laws of nature, governing the microstructure. (2) *P* is novel in that it is substantially different from its underlying microstructure. (3) The microstructure underlying *P* exhibits a greater degree of variance from moment to moment than does *P* (that is, *P* persists even if the microstructure fluctuates – even if the constituent parts of *P* change or are replaced). And (4) *P* causally influences at least some of the properties in the underlying microstructure.

The above characterization of emergent phenomena is borrowed from Robert L. Klee (1984). Emergence might be characterized in slightly different ways, but I think the above captures the essential emergentist argument. For another characterization, see Paul Humphreys' (1997) "Emergence, not supervenience." Humphreys defines emergence as follows: (1) emergent properties are novel, (2) emergent properties are qualitatively different from the properties from which they emerge, (3) an emergent property is such that it could not occur at a lower level, (4) different laws apply to emergent phenomena than to the microstructure such phenomena emerge from, (5) emergent properties "result from an essential interaction between their constituent properties, an interaction that is nomologically necessary for the existence of the emergent property" (Humphreys, 1997, p. S342), and (6) emergent properties are holistic (they are properties of the entire system and not just properties of the constitutive parts).

Bedau criticizes *strong* emergence on the grounds that it is metaphysically illegitimate and that it is inconsistent with materialism because it postulates downward causation (he also suggests that *strong* emergence is not scientifically useful). Weak emergence, on the other hand, Bedau claims, is: (i) metaphysically innocent because it does not commit itself to downward causation, (ii) scientifically useful in that the concept, as well as terminology, is ubiquitous in the thriving scientific research into complex systems, and (iii) it is consistent with materialism. Bedau further argues that weak emergence retains the two hallmarks of emergent properties.

Weak emergence faces at least two objections: (1) It can be objected that weak emergence applies too widely and too arbitrarily and thus that it is too weak (that is, weak emergence does not demark an interesting class of phenomena or it applies to phenomena that are not emergent). Bedau responds by stating that not all emergent phenomena need to be interesting to be labelled emergent. (2) One may also object that the concept of weak emergence is useless because we do not generally have proof that a given macrostate of a system is un-derivable without simulation. Bedau responds by stating that unproven weak emergence claims still possess substantial empirical support.

It may be useful to consider a possible reason for Bedau's claim that downward causation is metaphysically illegitimate. Jaegwon Kim (2000) argues that emergentism is committed to the supervenience thesis because "if two systems are wholly alike physically, we should expect the same mental properties to emerge, or fail to emerge, in each" (Kim, 2000, p. 38).

Let's turn to Broad's (1925) example of emergent phenomena in order to understand Kim's claim:

> Oxygen has certain properties and Hydrogen has certain other properties. They combine to form water, and the proportions in which they do this are fixed. Nothing that we know about Oxygen by itself or in its combinations with anything but Hydrogen would give us the least reason to suppose that it would combine with Hydrogen at all [and vice versa] … And most of the chemical and physical properties of water have no known connexion, either quantitative or qualitative, with those of Oxygen and Hydrogen. Here we have a clear instance of a case where, so far as we can tell, the properties of a whole composed of two constituents could not have been predicted from a knowledge of the properties of these constituents taken separately, or from this combined with a knowledge of the properties of other wholes which contain these constituents. (Broad, 1925, pp. 62-63)

It is important to note that even if the properties we associate with water are emergent, when oxygen and hydrogen combine to form $H_2O$, "the proportions in which they do this are fixed" (Broad, 1925, p. 62). That is, every time the proportions of hydrogen to oxygen exhibit this 2-1 ratio, we can expect the emergent properties to occur. Since emergent properties display this intimate connection to the structural properties of their constituent parts, they appear to be dependent on those micro-level structural properties of the lower-level parts that constitute the *emergent* whole. Emergentism, then, appears to be committed to the view that macro-level phenomena supervene on micro-level phenomena.

If emergent phenomena supervene on their micro-level constituents (and if the mind is an emergent property that supervenes on the physical properties of the brain), then the notion of downward causation (and thus also the notion of mental causation) faces the problem of causal exclusion. The problem of causal exclusion, as outlined by Kim, states that the physical cause threatens to exclude the mental cause. In other words, the micro-level cause threatens to exclude the macro-level cause (if there is such a thing at all). Assuming that mental event *M* supervenes on physical event *P* (as would be the case if mental properties were emergent properties), if *M* is said to cause a mental event *M**, then, if we adopt physicalism and reject any sort of dualism, *M** must also have a subvenient physical event *P** on which it (*M**) supervenes. If this is true, then since *M* is dependent for its existence on *P*, the occurrence of *M** must also be dependent on the existence of *P**. That is, if *P** did not exist, neither could *M**. Thus, since there must be a causal link between *P* and *P** at the lower, physical level, *P*'s causal efficacy appears to be sufficient for explaining the occurrence of *P**. If *M* supervenes on *P*, then *M*'s existence is possible if and only if *P* also exists (the same must be said of the relationship between *M** and *P**). This holds even if the multiple realizability thesis is true; even then, *M* will depend for its existence on some physical state (*M* must be realized by some *P*). Thus, it would appear that either *M* is not causally efficacious or that *M* can only be causally efficacious in the presence of *P*. If the first alternative is true, then *M* is merely epiphenomenal and if the latter is true (that *M* can only be causally efficacious if *P* is present), then we are faced with a case of over-determination (one reason why over-determination is problematic is that it actually contributes to the problem of causal exclusion). Either way, *M* does not seem to be necessary for *M** to obtain (all that is required is the presence of *P*). I cite Kim's much more eloquent explanation of this problem:

> [T]o acknowledge that *p* has also a physical cause, *p**, at [time] *t* is to invite the question: Given that *p* has a physical cause *p**, what causal work is left for *m* [a mental event] to contribute? The physical cause therefore threatens to exclude, and pre-empt, the mental cause. (Kim, 2000, p. 37)

Therefore, if mental events are emergent, but downward causation is problematic, then mental causation is problematic. It would appear that even though we have the experience of causing our actions (after all, we commonly perceive the causal efficacy of our own minds), our mental states may not actually be causally efficacious. If mental causation proves to be just an illusion (in light of the supposed illegitimacy of downward causation), then free will must also be illusory (see McDermott, 2001; Smilansky, 2000; and Wegner, 2002 for, respectively, computational, philosophical, and psychological explorations of the thesis that free will is an illusion).

If determinism proves to be true (and downward causation proves to be metaphysically illegitimate and therefore false), free will must be illusory because if it is not the case that we have the right kind of control over our actions, then we are still in possession of the experience of this kind of control. That is, we feel free when we act. If we are determined to act in the ways we do, then the experience of free will (the feeling of control) must be illusory.

Should we commit ourselves, then, to Bedau's *weak* emergent view and give up the prospect of metaphysical freedom (or at least abandon the possibility that emergentism is the key to libertarian freedom)? Perhaps emergentism may still offer some hope for libertarians. Strong emergence (especially in its more contemporary form) is weaker than the theories espoused by the British Emergentists and thus, may, in fact, prove to be just the kind of view for which the libertarian about free will is searching.

## 4. Macro-Determinism

The Nobel laureate neurobiologist and neuropsychologist R. W. Sperry (1986) claims that a strict micro-determinism cannot fully account for emergent properties. Sperry argues that emergent macro-level phenomena exert causal influence on their micro-level parts. Sperry explains, however, that such macro-determinism does not negate or displace micro-determinism, but rather that it supplements it. Thus, although Sperry's version of downward causation is reminiscent of the stronger versions of emergence, I read Sperry (and his supporters) as taking a middle ground between strong and weak emergence. In fact, Sperry's conception of macro-determinism is not as strong as that of the British Emergentists'. That is, Sperry's view does not disregard micro-determinism, but merely adds to it.

Sperry compares the mind exercising influence over the brain to a computer's software wielding downward causal control over the computer's micro-physical elements:

> [M]any agree that the 'macro' computer software programs exert downward causal control over their electronic and other micro-physical correlates and, just like the conscious mental programs

in the brain, have their own dynamics, properties, and laws of progression and interaction distinctly different from those of the underlying microphysics. (Sperry, 1986, p. 269)

Sperry explains that "[t]he emergent properties of the entirety and the laws for its causal interactions are determined by the spacing and timing of the parts as well as by the properties of the parts themselves" (Sperry, 1986, p. 266). Sperry, in his 1964 *James Arthur Lecture* titled "Problems outstanding in the evolution of brain function," gives a telling example of what he means by macro-determinism:

[R]ecall that a molecule in many respects is the master of its inner atoms and electrons. The latter are hauled and forced about in chemical interactions by the overall configurational properties of the whole molecule. At the same time, if our given molecule is itself part of a single-celled organism such as paramecium, it in turn is obliged, with all its parts and its partners, to follow along a trail of events in time and space determined largely by the extrinsic overall dynamics of *Paramecium caudatum*. (as cited in Sperry 1986, 265-266)

It would appear, then, that downward causation may, in fact, turn out not to be metaphysically illegitimate. However, the pressing question, one that Robert L. Klee (1984) considers, is whether macro-determinism is truly a case of downward causation or whether such apparent higher-level efficacy is in reality reducible to micro-level properties?

Klee argues for micro-determinism, the view that parts determine wholes. He states that it is "a prevalent belief among both scientists and philosophers that, for most systems in nature, higher-levels of organization are by and large determined (and therefore explained) by lower-levels of organization" (Klee, 1984, p. 44). Direct determination from the macro-levels to the micro-levels, according to Klee, seems somewhat mysterious and supporters of such views are forced to resort to "metaphors and analogies to illustrate the 'how' and the 'why' of this kind of macro-determinism … But is this a kind of determinative influence really in conflict with a micro-deterministic view of the world" (Klee, 1984, p. 60)? The metaphors and analogies Klee is referring to are Sperry's examples of a wheel and the molecules that constitute it and of a local eddy in a stream and the water molecules that make it up. He writes:

[t]he macro-determinative influence of consciousness on underlying micro-states of the brain is *like*, says Sperry, the macro-determinative influence that the whole wheel has on an individual molecule within it in virtue of which the molecule goes only where the whole wheel goes. Or again, says Sperry, it is like the kind of determinative influence a local eddy in a stream has on an individual water molecule within it in virtue of which the molecule goes in the direction in which the entire eddy is flowing. (Klee, 1984, p. 60)

Klee continues:

To be sure, this is the kind of determination which is 'holistic' insofar as the effect is transmitted to the molecule through a larger holistic micro-structure. But it does not seem ultimately to be a kind of determinative influence that can't be given a basically micro-level mechanism; that is, we want to know how the motion of the whole wheel influences the motion of the molecule and *that* story, while involving the holistic aspect of the molecule's participation in a larger structure, will primarily involve same level connections—micro-connections … The means of effecting the determinative influence seem same-level, i.e. a micro-story is to be told. (Klee, 1984, p. 61)

Klee's point is that the causal efficacy of any given system is to be found at the micro-level, even if we can engage in higher-level descriptions of systems interacting with other systems. The complex interactions of parts (both inter-systemic and intra-systemic) are ultimately micro-explainable. That is, all macro-level phenomena and events are reducible to their micro-level constituents and these constituents are ultimately responsible for the macro-level properties.

Timothy O'Connor (1994), in his defence of the possibility of macro-determinism, explains that Sperry's notion of downward causation is compatible with the supervenience thesis and with micro-determinism. He states that macro-determinism (as Sperry understands it) does not disrupt or intervene in the causal relations that occur at the micro-level, but rather that it supervenes in a way that leaves micro-interactions unaffected. O'Connor cites Sperry's (1991) paper "In defense of mentalism and emergent interaction":

These micro interactions and the interrelations of all the infrastructural components become embedded within, enveloped, and as a result are thereon moved and carried by the property dynamics of the larger overall system as a whole … that have their own irreducible higher-level forms of causal interaction. A molecule within a rolling wheel, for example, though retaining its usual inter-molecular relations within the wheel, is at the same time, from the standpoint of an outside observer, being carried through particular patterns in space and time determined by the over-all properties of the wheel as a whole. There need be no "reconfiguring" of molecules relative to each other *within the wheel itself*. However, *relative to the rest of the world* the result is a major "reconfiguring" of the space-time trajectories of all components in the wheel's infrastructure. (as cited in O'Connor, 1994, p. 101)

The idea is that the emergent property and the resulting relational structure of the micro-parts are instantiated simultaneously. That is, the relation between the macro and the micro is not that of an efficient cause to an effect, but rather that even though there is causal interaction at the micro-level, the macro-level dictates certain structural properties of the micro-parts and thus becomes a salient part of the causal story. O'Connor explains that on Sperry's view, "an emergent determines (in large part) a relational complex *that cannot be adequately described in terms of lower level components and their micro-relations* [and thus], the micro-physical laws are inadequate for, cannot be applied to, such situations" (O'Connor, 1994, p. 102). Sperry argues that "[m]icro-determinism is retained but is held to be incomplete, insufficient. The properties, forces and laws of micro-events are shown to be encompassed and superseded, not disrupted, by the properties, forces, and laws at macro-levels" (Sperry, 1986, p. 268). That is, macro-level determination does not displace, but only supplements micro-determination.

Thus, if we accept Sperry's macro-determinism and follow O'Connor in admitting that Sperry's approach is both conceptually and empirically viable, then perhaps free will (in the libertarian sense) is also conceptually and empirically viable. After all, it does seem to us that we are

free. The experience of free will accompanies many of our actions and we feel as though we could have done otherwise. That is, we feel as though we exert causal influence on our environments, bodies, and actions; we feel that we are self-determiners.

However, Klee's scepticism is not ungrounded. That is, even though wholes are structurally different from their constituents in that wholes are the structural complexes into which their constituent parts are arranged, it is these constituents that are ultimately responsible for the structural complexes of the wholes of which they are parts. Whatever emergents there may be (and however "autonomous"), they necessarily depend on their micro-structure and thus it is the micro-structure that is responsible for any macro-level properties and macro-level causation. But if this is the case, then we either run into the problem of over-determination or the problem of causal exclusion. The principle of parsimony suggests that since micro-determinism appears to give us a complete explanatory story, we need not postulate macro-determinism.

It may be worth examining Sperry's examples in order to better understand Klee's concern. Sperry states that a whole wheel (or a local eddy in a stream) has macro-determinative influence on the individual molecules that make it up (the molecules that make up the wheel or the eddy). However, as Klee explains, the "higher-level motion of the whole wheel or eddy has an influence on the lower-level individual molecule through the fact that the molecule bears intimate structural micro-connections with neighboring molecules in the wheel or eddy" (Klee, 1984, pp. 60-61). The wheel or the eddy do not set the molecule into motion, but rather the eddy's or the wheel's motions are a result of the molecule's intimate connection to other molecules. The eddy or the wheel are constituted by certain molecules that are in a certain relational configuration. The motion of the molecule, then, "is dictated to the molecule in virtue of its participation in the total micro-structure of the wheel or eddy" (Klee, 1984, p. 61), but the micro-structure that exerts a causal influence on the molecule is itself a lower-level phenomenon.

One way of understanding Klee's objection is by means of an analogy with Conway's Game of Life. The Game of Life is a cellular automaton devised by the British mathematician John Conway. The universe of the Game of Life consists of a two-dimensional grid of square cells or pixels. Each cell can be in one of two possible states (either *live* or *dead*, ON or OFF). Time, in the Game of Life, is discrete and the state of every cell at time $t$ is a function of the states of its neighbouring cells at time $t_1$. The micro-dynamic governing the states of every cell (at any given time step) consist of four rules (dubbed the "birth-death rule"):

1. Any live cell with fewer than two live neighbours dies, as if by loneliness.
2. Any live cell with more than three live neighbours dies, as if by overcrowding.
3. Any live cell with two or three live neighbours lives, unchanged, to the next generation.
4. Any dead cell with exactly three live neighbours comes to life.

There are various stable patterns that "emerge" in the Game of Life, some of the favourites are given names (e.g. gliders, which maintain their structural integrity as the pixel pattern is transcribed across the grid in accordance with the four simple rules mentioned above). Patterns are also grouped into categories (e.g. still lifes, oscillators, spaceships, etc.).

If we identify the molecules in Sperry's wheel or eddy example as being analogous with the single cells or pixels in the Life World and if we draw an analogy between the wheel or eddy and the *emergent* patterns that are constituted by the pixels and the micro-dynamic, then Klee's case should become clear. The pixels or cells enter into a certain micro-structure when they interact with each other. Their interaction happens in virtue of the birth-death rule. A glider in the Life World, then, can be imagined to exert a causal influence on the pixels that constitute it because the pixels appear to move in the direction the glider takes them inside the two-dimensional Life World space. However, the existence of the glider itself depends, first and foremost, on the life-death rule's influence on the pixels. The pattern (i.e. the glider), is an emergent (though only a weakly emergent) and persisting entity, but is fully caused by the micro-dynamic that influences the pixels, which constitute the pattern.

Returning to Sperry's example, the molecules in the eddy (or the wheel) do go wherever the eddy takes them, but the eddy itself does not govern their movement because the eddy's movement is, in fact, nothing more than the result of the complex interaction between the molecules that constitute it. And the molecules themselves are governed by micro-level forces. Ultimately, then, the only explanatory causal story we need to give is a micro-level story. Perhaps the only coherent account of emergence is Bedau's weak emergence (however *weak* it may actually be).

If downward causation is not a feature of emergent phenomena, then emergent entities (such as minds, for instance) cannot exercise causal influence on their constituents and are themselves fully determined by their micro-level parts. It would appear, then, that in the absence of downward causation (which seems to be implausible at best), libertarians about free will must either find another metaphysical (or scientific) grounding for their conception of freedom or they must give it up in favour of a hard-deterministic position (or alternatively a compatibilist view). In either case, the notion of a self-determining agent faced with open alternative possibilities between which the agent "freely" chooses must be abandoned (although open alternative possibilities would actually exist on an indeterministic view, that would not necessarily give rise to the type of control required for free will – chancy-ness or randomness does not lend itself to genuine metaphysical freedom). Roderick Chisholm's (1964) Aristotelian unmoved mover does not make much sense in a world ultimately governed by the *mover's* parts (or even, as seems to be the case given the discussion thus far, governed by the parts of the mover's parts).

## 5. A World Without Freedom: A World Without Macro-Determinism

If the strong emergentist conception of macro-determinism is untenable, then it does not appear to be the case that libertarian free will is possible (even if we assume that the

types of weakly emergent phenomena Bedau discusses do actually exist). What are we left with if macro-determinism and thus, by extension, the libertarian conception of free will prove to be false? We may not be losing out on that much (or perhaps even on nothing at all) if determinism proves to be true while libertarian free will turns out to be a mere illusion. As Dennett explains:

> After all, if determinism is true now, it always has been true … Modern science isn't *making* determinism true, even if it is discovering this fact, so things aren't going to get worse, unless it is believing in determinism rather than determinism itself that creates the catastrophe. (Dennett, 1984, p. 15)

Assuming that we all have the experience of free will (that is, assuming that we all feel as though we act freely and as though we exercise an agent-causal control over ourselves and our actions), if determinism is the case, then the experience of free will must be just an illusion. That is, if we grant the truth of determinism (and thus that metaphysical freedom does not exist), we must be illusionists about free will because we are all "stuck," as it were, (at least for the most part) with an experience of freedom. The same would be true, however, if the universe turns out to be indeterministic. Peter van Inwagen (1998) raises this worry. He writes:

> But then, if the world is indeterministic, isn't it just a matter of chance how things *did* happen in the one, actual course of events? And if what we do is just a matter of chance — well, who would want to call that freedom? (van Inwagen, 1998, p. 370)

Although Dennett's words may, at first glance, be somewhat comforting (after all, if determinism is true, nothing will really change for us since it is true now and always has been the case), there appears to be a problem with reconciling determinism (or, alternatively, indeterminism) and moral responsibility. That is, since our commonsense notion of moral responsibility depends on our belief in our own free agency, if we abandon libertarianism, then what can we make of moral responsibility? It may, in fact, be the case that even though the truth of determinism is quite harmless, the knowledge of (and thus the belief in) the truth of determinism, as Dennett suggests, can cause the catastrophe. There have been many attempts at reconciling moral responsibility with determinism. I do not have the luxury of space, in this paper, to sketch even a few of these approaches and thus, I shall focus on just one (arguably the most robust). See Lenart (2007) for an initial sketch of the following argument

## 6. Strawson's Reactive Attitudes

Peter Strawson (1963/2004) in his seminal paper "Freedom and resentment," offers an interesting solution to the problem of moral responsibility. He argues that punishment, ethical approval, or moral condemnation are practices or attitudes central to human relationships and interactions. He mentions two types of attitudes: (1) attitudes that permit (and sometimes even require) detachment from an action or from the agent responsible for an action, and (2) "non-detached attitudes and reactions of people directly involved in transactions with each other … attitudes and reactions of offended parties and beneficiaries … [these being] such things as gratitude, resentment, forgiveness, love, and hurt feelings" (Strawson, 1963/2004, p. 75).

Interpersonal relating gives rise to these reactive attitudes because it genuinely matters to us whether people are affectionate toward us, or if they exhibit contempt or malevolence, etc. While Strawson focuses on cases where reactive attitudes are natural, he also reflects on cases where special considerations might modify or mollify the natural feelings of affection or resentment; these include such things as unintentional actions or instances where an agent could not have done otherwise due to direct coercion. Nevertheless, these kinds of circumstances are not generalizable and thus under normal conditions, we do not usually, nor are we expected to, suspend our ordinary reactive attitudes. Such circumstantial excuses, however, "do not invite us to view the *agent* as one in respect of whom these attitudes are in any way inappropriate" (Strawson, 1963/2004, p. 77). A second set of excusing conditions consists of (1) individuals temporarily acting out of character, and (2) individuals permanently incapable of appropriately engaging in interpersonal relationships:

> The second and more important subgroup of cases allows that the circumstances were normal, but presents the agent as psychologically abnormal—or as morally undeveloped. The agent was himself; but he is warped or deranged, neurotic or just a child. When we see someone in such a light as this, all our reactive attitudes tend to be profoundly modified. (Strawson, 1963/2004, p. 79)

Strawson argues that human beings are naturally committed to interpersonal human relationships, which require interpersonal attitudes (which are, simply stated, attitudes we adopt or have towards others). Interpersonal attitudes, in turn, require reactive attitudes, which are natural human reactions to the actions of others. He writes: "This commitment is part of the general framework of human life, not something that can come up for review … [hence] the truth or falsity of a general thesis of determinism would not bear on the rationality of *this* choice" (Strawson, 1963/2004, p. 83).

The reactive attitudes can be of a vicarious nature (where we experience reactive attitudes on behalf of others). Such reactive attitudes find their analogues in our morality as moral reactive attitudes, which are of a vicarious nature. Such vicarious reactive attitudes are sympathetic, impersonal, disinterested, or generalized analogues of the reactive attitudes and deal not so much with resentment as with moral indignation or disapprobation. "They are reactions to the qualities of others' wills, not towards ourselves, but towards others. Because of this impersonal or vicarious character, we give them different names" (Strawson, 1963/2004, p. 83), we call them moral reactive attitudes. Strawson emphasizes: "It is not that these attitudes are essentially vicarious—one can feel indignation on one's own account—but that they are essentially capable of being vicarious" (Strawson, 1963/2004, p. 84). Moreover, they can be applied to ourselves (that is, they can serve as demands on ourselves for others). Such vicarious and self-directed reactive attitudes find their analogues in our morality as moral reactive attitudes.

Strawson argues that we are incapable of suspending our reactive attitudes because our human nature (i.e. our social nature) requires interpersonal relationships. Thus, the reactive attitudes and their moral analogues are as deeply ingrained in our nature as is our inclination toward interpersonal relationships. Strawson writes:

> Finally, to the further question whether it would not be *rational*, given a general theoretical conviction of the truth of determinism, so to change our world that in it all these attitudes were wholly suspended, I must answer, as before, that one who presses this question has wholly failed to grasp the import of the preceding answer, the nature of the human commitment that is here involved: it is *useless* to ask whether it would not be rational for us to do what is not in our nature to (be able to) do. (Strawson, 1963/2004, p. 87).

## 7. Illusionism About Free Will

Many of our choices are accompanied by the phenomenal experience of what I have been calling metaphysical freedom; that is, many of our choices are coupled with the sensation that *we could have done otherwise* and with the feeling that *it was one's will that caused one's action*. In his book, *The Illusion of Conscious Will*, the Harvard psychologist Daniel Wegner writes:

> The notion that will is a force residing in a person results in a … problem. Hume … pointed out that causality is not a property inhering in objects … you can't *see* causation in something but must only infer it from the constant relation between cause and effect. Every time the ball rolls into the pins, they bounce away. Ergo, the ball caused the pins to move. But there is no property of causality … hanging somewhere in space between the ball and pins … Causation is an event, not a thing or a characteristic or attribute of an object. In the same sense, causation can't be a property of a person's conscious intention. You can't *see* your conscious intention causing an action but can only infer this from the constant relation between intention and action. (Wegner, 2002, p. 13)

Wegner defines conscious will as (1) the phenomenal experience of performing an action (actions either feel willed or they do not), or (2) as the causal link between mental states and actions. The mistaken assumption that (1) and (2) are identical is, according to Wegner, precisely the source of the illusion of conscious will.

Wegner's Theory of Apparent Mental Causation states that "[*p*]*eople experience conscious will when they interpret their own thought as the cause of their action*" (Wegner, 2002, p. 64). The phenomenal experience of consciously willing something is produced by what Wegner calls *priority*, *consistency*, and *exclusivity*. "For the perception of apparent mental causation, the thought should occur before the action, be consistent with the action, and not be accompanied by other potential causes" (Wegner, 2002, p. 69).

The Priority Principle is a causal principle stating the precedence of causes to their effects. If a person experiences X as causing Y, then, according to the Priority Principle, X must necessarily precede Y, and X cannot occur after Y, nor too long before the occurrence of Y. For example:

> When one billiard ball strikes another, the struck ball moves in the same general direction that the striking ball was moving. We do not perceive causality very readily if the second ball squirts off like squeezed soap in a direction that, by the laws of physics, is inconsistent with the movement of the first ball. (Wegner, 2002, p. 78)

Wegner explains the Exclusivity Principle as follows: "[w]hen their own thoughts do not appear to be the exclusive cause of their action, they experience less conscious will. And when other plausible causes are less salient, in turn, they experience more conscious will" (Wegner, 2002, p. 90). Wegner postulates that the causes of human actions are in fact complex mechanisms hidden from conscious observation. He writes: "[w]e must remember that this analysis suggests that the real causal mechanisms underlying behavior are never present in consciousness. Rather, the engines of causation operate without revealing themselves to us and so may be unconscious mechanisms of mind" (Wegner, 2002, p. 97). He further argues that actions can occur without sufficient intentions, and that we compensate for such moments by confabulating the non-existent intentions.

> When life creates all the inevitable situations in which we find ourselves acting without appropriate prior conscious thoughts, we must protect that illusion of conscious will by trying to make sense of our action. We invent relevant thoughts according to the template that conscious agency suggests. (Wegner, 2002, p. 157)

He adds that cognitive dissonance explains how people revise their attitudes in order to justify their actions.

> In a nutshell, the theory says this happens because people are motivated to avoid having their thoughts in a dissonant relationship, and they feel uncomfortable when dissonance occurs. The strongest dissonance arises when a person does something that is inconsistent with a preexisting attitude or desire. (Wegner, 2002, p. 172)

Wegner compares the will to a compass, explaining that in the same way that a compass reading does not have any causal efficacy on the ship's actual direction, the will itself does not cause human behaviour, but, like a compass reading, the will can be a good gauge to which we can refer as we steer, internalize, and appropriate our actions. He writes: "the occurrence of conscious will brands the act deeply, associating the act with self through feeling, and so renders the act one's own in a personal and memorable way. Will is a kind of authorship emotion" (Wegner, 2002, p. 325).

> Conscious will is particularly useful, then, as a guide to ourselves. It tells us what events around us seem to be attributable to our authorship. This allows us to develop a sense of who we are and are not. It also allows us to set aside our achievements from the things that we cannot do. And perhaps most important for the sake of the operation of society, the sense of conscious will also allows us to maintain the sense of responsibility for our actions that serves as a basis for morality. (Wegner, 2002, p. 328)

## 8. Reactive Attitudes and the Benefits of Illusionism

Saul Smilansky observes that "if libertarian assumptions *carry on their back* the CC [Core Conception] distinctions,

which would not be adhered to sufficiently without them, an illusion which defends these libertarian assumptions seems to be just what we need" (Smilansky, 2000, p. 173) (the Core Conception is the elementary ethical conception that takes as its focus the necessity of considering free will as a prerequisite for morality). In other words, the belief in free will may serve as a vehicle for our acting morally much of the time.

In "Free will and respect for persons," Smilansky argues that the truth or falsity of determinism does not affect our ability to hold people morally accountable for their actions. As already stated above, Strawson argues for a similar conclusion. It is crucial, however, that morally accountable individuals be in possession of certain rational capacities (such as the capacities for awareness, deliberation, choice, and intentional action), which enable them to act *responsibly*. On the Strawsonian view, "normal" adults belong to a *community of responsibility* (which, roughly stated, consists of members of a certain community who share a given set of moral reactive attitudes). It is quite irrelevant, for both Strawson and Smilansky, whether a *normal* adult human has metaphysical freedom (libertarian free will) or whether she is fully determined. What is of importance is that she has all the required capacities to be held responsible by her peers in the "Community of Responsibility." Smilansky, however, continuously (both in his 2000 book as well as in his 2005 article and on many occasions in each) worries about the problem of the ultimate arbitrariness of all moral judgments (this ultimate arbitrariness stems from the admission that free will is nothing more than an illusion):

> While membership in a Community of Responsibility permits punishment of the guilty student, it at the same time forbids 'punishment' of the innocent one. Nevertheless, the actions of the drug dealer [the guilty student] were, in one way, merely an unfolding of the given, of matters that, causally constituting her, were ultimately beyond her control. Together with the moral obligation to respect and to track (in our own reactions and practices) identity, choice, and responsibility, we must also not forget the ultimate *arbitrariness* of it all. (Smilansky, 2005, p. 256)

Nonetheless, Smilansky argues that Strawsonians should adopt illusionism because a state of affairs where we would not need recourse to illusion, but determinism would still be true, carries a price we cannot afford. The knowledge of determinism (and thus the knowledge of our lack of control over ourselves and our actions), according to Smilansky, would "put our moral house at grave risk" (Smilansky, 2005, p. 257). He continues: "The moral house we have is essentially a Community of Responsibility … In short, the ethical importance of the Community of Responsibility should be taken very seriously, but the ultimate perspective threatens to *present* it as a farce, a mere game without foundation" (Smilansky, 2005, p. 257). Therefore, the *illusion* of freedom is actually both necessary and good for the Strawsonian notion of reactive attitudes. Regarding the necessity of the illusion of metaphysical freedom, Smilansky writes:

> Respect for persons requires on the one hand respect for agency, the establishment of a moral order based on responsibility, and the attempt at human empowerment within compatibilist spheres; on the other hand, it requires recognition of the limitations and shallowness of these spheres, where everything that goes on is ultimately an unfolding of the given, beyond anyone's control. This dissonance already calls for illusion to serve a 'functional' role, that of safeguarding the partly valid compatibilist-level 'form of life' (a primary condition for respect for persons) from the threat of the ultimate hard determinist perspective that levels all of us. But beyond the 'functional' stage lies the 'existential' stage, where philosophically we can recognize how intimately our fundamental evaluations of ourselves and of others, and of our reactions to one another, depend on the false libertarian picture. We confront the deep dangers of awareness and internalization of the truth. At the depths, the libertarian illusion is constitutive of our very humanity; it is a condition for deep self-respect and for respect for persons. (Smilansky, 2005, pp. 260-261)

Even if the emergentist view does not offer libertarian freedom, our moral lives appear to be safe from the dangers posed by determinism because, for whatever evolutionary reason, we have been endowed with the illusion of free will, which acts as a vehicle for human dignity and our commonsense notion of moral responsibility.

Moreover, the mere illusion of conscious willing and metaphysical freedom has its advantages. Wegner reports on findings by Rodin & Langer (1976 and 1977) that elderly individuals who are given new responsibilities like watering plants, which generally amount to novel control opportunities in their lives, tend to display renewed psychological and physical resilience. Along these same lines of discovery, Bulman & Wortman (1977) found that victims of psychologically debilitating accidents who assumed responsibility for their misfortune were better able to cope with their calamity. Wegner comments: "the habit of taking responsibility seemed to carry over from the accident into the pursuit of adjustment in the aftermath … it is reasonable for a person who perceives control in one area to suspect the possibility of such control in another" (Wegner, 2002, p. 330).

Wegner argues that the illusion of metaphysical freedom is the building block of both human psychology and the very fabric of our social life. "It is only with the feeling of conscious will that we can begin to solve the problems of knowing who we are as individuals, of discerning what we can and cannot do, and of judging ourselves morally right or wrong for what we have done" (Wegner, 2002, p. 342). The illusion of free will, according to Wegner, is the source of our humanity and the dignity with which we endow it.

## 9. Weak Emergence, Illusionism, and Human Dignity

The dignity with which we endow human beings is a measure of the respect bestowed upon them. Dignity can be understood as a relation of respect toward the moral interests of individuals. This can certainly be unpacked in a number of different ways, but the Strawsonian model of reactive attitudes is definitely robust enough to encompass and bestow this sense of dignity through the processes involved in what Strawson calls vicarious reactive attitudes (see Wilson-Lenart (2014) for a care-ethically grounded discussion of dignity).

Insofar as illusionism about free will grounds the Strawsonian notion of reactive attitudes, so the illusion of freedom also grounds the mechanisms that bestow dignity on moral actors, agents, and patients. This particular connection between illusionism and dignity may, however, seem problematic to those with reductive tendencies. It may be argued, after all, that since human agency is ultimately reducible to the motions of our constitutive parts, not only human freedom, but also human morality and dignity are similarly reducible to nothing more than simple, predictable processes, which themselves are neither moral nor dignified.

While the postulation of downward causation would perhaps make for the strongest response to this objection, emergentism in its weaker forms nevertheless offers a non-reductive picture of human beings and their complex mental lives. Human beings (or more technically, *persons*), on this view, are wholes that are irreducible to their constitutive parts (the parts that constitute the individualized physical instantiations of the species *Homo sapiens*); and while the absence of downward causation most likely entails that human beings do not enjoy metaphysical freedom, persons nevertheless navigate a complex moral matrix constituted by dynamic and emergent moral reactive attitudes, and in virtue of so doing, persons generate dignity in a way that machines or naturalistic processes cannot. Moreover, the persistent and unshakable illusion of freedom, which buttresses these moral reactive attitudes (and thus the resultant dignity of entities with such moral capacities), is itself an emergent manifestation of the complex mental properties that generate it, and in so being, is an irreducible property of human minds and human nature, one that not only makes human beings special and unique, but also morally considerable and dignified.

## 10. Conclusion

While Sperry's emergentism does not ultimately succeed in promoting a notion of downward causation of sufficient utility to libertarians in search of metaphysical freedom, it nevertheless does contribute to an understanding of the human agent as a unique and dignified whole that is not reducible to its simpler, morally irrelevant parts. This view of the human agent is further buttressed by both the philosophical and psychological arguments offered, respectively, by Smilansky and Wegner, who argue that the very processes responsible for the illusion of conscious willing and metaphysical freedom are in fact also responsible for both the overall well-being as well as dignity of individuals (complex systems) with the capacity to generate the phenomenal experience of freedom.

Furthermore, although given all of the above considerations, conscious willing and thus the phenomenal experience of metaphysical freedom are likely merely illusory, beings like us (ones that emerge out of the complex whirl of various lower-level processes) are nevertheless not morally inert, as Starwson's proposal that reactive attitudes are capable of grounding the moral fabric of interpersonal relating helps alleviate the libertarian's worry about moral responsibility.

## Works cited


Bedau, M. A. (2008). Weak emergence. *Noûs*, *31*. 375-399. https://doi.org/10.1111/0029-4624.31.s11.17

Broad, C. D. (1925). *Mind and its place in nature*. New Harcourt, Brace & company, Inc.

Bulman, R. J., & Wortman, C. B. (1977). Attributions of blame and coping in the "real world": Severe accident victims react to their lot. *Journal of Personality and Social Psychology*, *35*(5). 351-363. https:// doi.org/ 10.1037/0022-3514.35.5.351

Chisholm, R. M. (1964). *Human freedom and the self*. University of Kansas.

Crane, T. (2003). *The mechanical mind: A philosophical introduction to minds, machines and mental representation*. 2nd Ed. Routledge.

Dennett, D. C. (1984). *Elbow room: The varieties of free will worth wanting*. The MIT Press.

Flanagan, O. (2002). *The problem of the soul: Two visions of mind and how to reconcile them*. Basic Books.

Gribbin, J., Gribbin, M. (2001). *Stardust: Supernovae and life, the cosmic connection*. Yale University Press.

Humphreys, P. (1997). Emergence, not supervenience. *Philosophy of Science*, *64*(4). Proceedings of the 1996 Biennial Meetings of the Philosophy of Science Association. Part II: Symposia Papers. S337-S345.

Kim, J. (2000). *Mind in a physical world: An essay on the mind-body problem and mental causation*. MIT Press.

Klee, R. L. (1984). Micro-determinism and concepts of emergence. *Philosophy of Science*, *51*(1). 44-63.

Lenart, B. A. (2007). *Hardwired freedom: Illusion as a vehicle for moral responsibility* [Unpublished master's thesis]. University of Alberta. https:// era.library.ualberta.ca/ items/ 3abc4c40- 0b0d- 4f83- 9ff9-9f6b0 f26f59a#.WKXmRhIrI1h

Lenart, B. A. (2022). Metaphysical compatibilism and the ontology of trans-world personhood: A neo-Lewisian argument for the compatibility of divine foreknowledge (determinism) and metaphysical free will. *Metaphysica: International Journal for Ontology & Metaphysics*, 23(2). 1-23. https://doi.org/10.1515/mp-2021-0041

McDermott, Drew V. (2001). *Mind and mechanism*. MIT Press.

Nagel, T. (1976/1979). Moral luck. In Thomas Nagel (Ed.), *Mortal Questions* (pp. 24-38). Cambridge University Press. 1979. 24-38.

Nozick, R. (1981). *Philosophical explanations*. Harvard University Press.

O'Connor, T. (1994). Emergent properties. *American Philosophical Quarterly*, *31*(2). 91-104.

Popper, K. R. (1950). Indeterminism in quantum physics and in classical physics. Part I. *The British Journal for the Philosophy of Science*, *1*(2). 117-133.

Rodin, J., & Langer, E. J. (1976). The effects of choice and enhanced personal responsibility for the aged: A field experiment in an institutional setting. *Journal of Personality and Social Psychology*, *34*(2). 191-198.

Rodin, J., & Langer, E. J. (1977). Long-term effects of a control-relevant intervention with the institutionalized aged. *Journal of personality and social psychology*, *35*(12). 897-902. https:// doi.org/ 10.1037//0022-3514.35.12.897

Smilansky, S. (2005). Free will and respect for persons. *Midwest Studies in Philosophy*, *29*(1). 248-261. https:// doi.org/ 10.1111/ j.1475-4975.2005.00115.x

Smilansky, S. (2000). *Free will and illusion*. Clarendon Press.

Sperry, R. W. (1991). In defense of mentalism and emergent interaction. *Journal of Mind and Behavior*, *12*. 221-245.

Sperry, R. W. (1986). Macro- versus micro-determinism. *Philosophy of Science*, *53*(2). 265-270.

Sperry, R. W. (1964). Problems outstanding in the evolution of brain function. *James Arthur Lecture*. American Museum of Natural History.

Strawson, P. (1963/2004). "Freedom and Resentment." In Gary Watson (Ed.), *Oxford Readings in Philosophy: Free Will, 2nd Ed*. (pp. 72-93). Oxford University Press.

van Inwagen, P. (1998). The mystery of metaphysical freedom. In Peter van Inwagen and Dean W. Zimmerman (Eds.), *Metaphysics: The Big Questions* (pp. 365-374). Blackwell Publishing.

van Inwagen, P. (1975/2001). The incompatibility of free will and determinism. In Laura Waddell Ekstorm (Ed.), *Agency and Responsibility: Essays on the Metaphysics of Freedom* (pp. 17-29). Westview Press. 2001.

Wegner, D. M. (2002). *The illusion of conscious will*. MIT Press.

Wilson, R.A. and Lenart, B. A. (2014). "Extended Mind and Identity". In Jens Clausen & Neil Levy (eds.), *Handbook of Neuroethics*, Springer. pp. 423-439.