Buddhist Meditation and the Possibility of Free Will

Abstract

I argue that an analysis of Buddhist meditation theory and practice may be used to ground a model of the possibility of free agency that stands up against four powerful arguments for free will skepticism in contemporary analytic philosophy: Peter van Inwagen's consequence argument, which asserts that if choices are lawfully necessary consequences of prior events, then they are unfree; Derk Pereboom's two arguments for hard incompatibilism: the manipulation argument, which asserts that manipulated choices are unfree, determinism is functionally equivalent to manipulation, and thus determined choices are unfree; and the randomness argument, which asserts that we cannot claim authorship over random neural events; and Galen Strawson's impossibility argument, which asserts that choices are always conditioned by mental states, so unconditioned free will is impossible. Although Buddhism sees the entire process that begins with beliefs and desires and proceeds through deliberations, decisions, act-initiations, and efforts into actions as an ultimately impersonal or agentless process, Buddhism is nonetheless capable of formulating the diametrical opposite of Strawson's impossibilism and Pereboom's hard incompatibilism, what I call *possibilism* or *soft compatibilism*, the view that free choices and actions can emerge from conditioned or unconditioned mental states, despite whether or not the world is deterministic, and independently of such considerations. None of this should be taken to suggest that Buddhism, as is, contains or endorses a theory of free will. That is not at issue here. The point here is that Buddhism may formulate such a theory.

Powerful arguments against the very possibility of free will have become dominant in contemporary Western analytic philosophy. I argue that Buddhism has resources that may be deployed to formulate a position that may stand up against four of the most powerful skeptical arguments against belief in free will. I support this claim, roughly, in two very general stages. First, I outline most of the basic terminology and major positions that define the free will debate, and then I review the four major arguments that challenge belief in free will. And second, I then show how Buddhist ideas may be used to critique those four skeptical arguments. Each of these two general stages of the discussion will have several parts, some of which will not appear in a simple sequence, but all of which are arranged for progressively complex comprehension.

Major Terminology, Positions, and Arguments on Free Will

I will summarize the main terminology, positions, and arguments on free will here. However, for an excellent, concise, but thorough introduction to the free will issue that conceptually simplifies all the complexities for readers new to the free will issue and that narrows them all down to just one open scientific question (about whether or not our 'torn' decisions – those decisions about which we are genuinely undecided and that matter to us – are neurally determined), see Balaguer (2014); see Raz (2013), chapter 7, for a shorter, equally clear analysis of the issue.

Autonomists believe that sometimes humans exhibit an amount or type of free will that is sufficient for purposes of holding them morally responsible for their actions. Some autonomists think that for free will to obtain we cannot live in a deterministic world, for in such a world every world-state is uniquely and necessarily determined by the previous world-state in accordance with immutable laws of nature, in such a way that no other world-state other than the actual world-state is logically possible, and thus no human being could ever bring about any world-state other than the world-state that was determined by the previous world-state, in which case there can be no free will in a deterministic world. These autonomists are *incompatibilists* of a certain type, *libertarians*: they think free will exists but is incompatible with *determinism*, the view that every event or world-state (the sum total state of all events constituting the world at an instant) is necessitated by the previous event or world-state, respectively, in accordance with all the laws of nature (physics, chemistry, and so forth). Other autonomists think we can be autonomous even in a deterministic world, under the right conditions, say, if our choices are determined to be rational, reason-responsive, voluntary, unconstrained, and so forth; these are *compatibilists*.

Compatibilists consider determinism, the doctrine of universal lawful causation, logically consistent with free will in the moral-responsibility-entailing sense; those compatibilists who also think determinism is true are *soft determinists*. Incompatibilists who think determinism is true and thus that we lack free will are *hard determinists*. Some think free will is not only incompatible with determinism (they are hard determinists), but they also think free will is with *indeterminism*, the view that determinism is false (they are therefore *hard indeterminists*); because they combine hard determinism and hard indeterminism, they are *hard incompatibilists*. *Impossibilists* are another form of *free will skeptics*, that is, philosophers who think free will is impossible (like hard incompatibilists), but they think it is impossible independently of whether determinism is true or false; that is, they differ from hard incompatibilists in that hard incompatibilists think there can be no free will because determinism rules it out and so does indeterminism, whereas impossibilists think free will is impossible on other grounds having nothing to do with the truth or falsity of determinism, but simply because every choice is a function of, or conditioned by, the mental state that is occurent at the moment of choice.

I formulate a position opposite of impossibilism and/or hard incompatibilism, which I describe as *possibilism* or *soft compatibilism*, by contrast. Possibilists think free will is possible regardless of what mental state one might be in, contrary to Strawson, and soft compatibilists think free will is compatible with the world being deterministic and it is also compatible with the world being indeterministic. Just as impossibilists and hard incompatibilists may be grouped together as free will skeptics, *autonomy skeptics*, *pessimists*, or just *non-autonomists*, possibilists and soft compatibilists may be grouped together as *free will gnostics*, *autonomy gnostics*, *optimists*, or just *autonomists*.

One of the most powerful and seminal incompatibilist arguments is van Inwagen's *consequence argument*, which asserts, basically, that if determinism is true, then agents can never do other than what they are determined to do, and thus they cannot be free. Although van Inwagen argues here only for incompatibilism, hard determinists simply add to the incompatibilist's consequence argument a premise to the effect that determinism is true at the non-quantum level (based on the bulk of non-quantum-level science), and thus they conclude that there is no free will. Libertarians are autonomists who, conversely, simply add to the incompatibilist's consequence argument a premise to the effect that we do possess free will for various reasons (such as experience, phenomenology, and its necessity for morals and other basic features of our humanistic self-conception), in which case they conclude that determinism must be false, at least of some central feature of human agency, even if it is true of everything else.

Another one of the most powerful skeptical arguments against free will is Derk Pereboom's *manipulation* or *four-case argument* against compatibilism. This argument sets forth a first case of manipulation in which it is allegedly clear that an agent who is secretly neurally manipulated is not morally responsible, and then proceeds through three other cases, the last of which involves only the claim that the world is deterministic but involves no manipulation at all, but alleges that the agent is also not morally responsible, and challenges the compatibilist to identify a morally relevant distinction between any features of any two adjacent cases, claiming it is impossible to do so. Pereboom more broadly advocates hard incompatibilism, which simply adds to the manipulation argument for hard determinism the hard indeterminist's *randomness argument* to the effect that if indeterminism is true, then agents' choices are random and thus not

truly "up to" them, in which case free will is incompatible both with determinism and with indeterminism.

Perhaps the most powerful skeptical argument against free will is Galen Strawson's *impossibility argument*. Strawson argues, briefly, that free will is impossible – independently of considerations about whether determinism is true or false – because every choice we ever make is necessarily a function of, and thus conditioned by, whatever mental state we are in at and just prior to the moment of choice, something Buddhism would generally accept, but, Strawson adds, we cannot be completely and ultimately responsible for the mental state we happen to be in at any given moment, and so no choice is totally free or unconditioned. Since we cannot create ourselves *ex nihilo*, we can never be in a mental state that was not conditioned by the previous mental state, and so on, back to our earliest mental state(s). But if we can never be in a mental state that is unconditioned, we can never be ultimately responsible for any mental state we are in, nor for any choice we make that is a function of, or conditioned by, that mental state. Thus, free will in the moral-responsibility-entailing sense is impossible.

Compatibilism-Relevant Ideas

Let us begin with a discussion of some ideas in Buddhism that will be relevant to the issue of free will and its compatibility with determinism, indeterminism, and the conditioning of our mental states, that is, Buddhist ideas that are supportive of autonomism. I argue that meditation and related volition-regulating disciplines that lead Buddhists toward liberation increase free-will-related abilities in the soft compatibilist sense, and that an analysis of these abilities thus provides the basis for a Buddhist-inspired counter to all of these skeptical arguments.

The Buddhist meditative path is designed to eliminate what Buddhist theory sees as the illusory belief in an ontologically substantive ego or self, and along with it the unregulated expression of ego-volitional preferences through spontaneous or deliberated actions, which unregulated expression is seen as fueling the illusory sense of self or ego. However, the unregulated expression of ego-based volitions is, from a Western perspective, intuitively a significant part of what is thought to be constitutive of free will. Western thought prizes the free-will-related ability to act on one's desires and preferences, to do as one pleases, and to make choices that are up to oneself whenever considered alternatives pull at one in more than one direction. Thus, whereas Buddhism eschews the idea of catering to ego-volitional preferences altogether, Western philosophy has been hard-pressed to square these free-will-related abilities with determinism. In earlier times, it was hard-pressed to square these abilities with God's foreknowledge, and more recently it is hard-pressed to square them even with indeterminism.

The closest to this that Buddhism has been hard-pressed to explain is the problem of how karmic volitional formations can accrue to individuals that in ultimate reality are thought to lack ontologically substantive selves, or how reincarnation is possible in the absence of a real self. Historically, Buddhism has not been concerned with how determined individuals may be morally responsible agents. That is because Buddhism sees agents as ultimately impersonal, and it sees the illusion of agency or selfhood as the key cause of suffering, and thus it is more concerned with eliminating ego-volitional agency altogether in favor of selfless/altruistic volition than it is interested in defending it. Enlightenment, for Buddhists, consists in the permanent realization of selflessness, both cognitively and volitionally.

Likely due to its recent encounters with Western philosophy and science, however, scholars of Buddhism have recently tried to articulate what Buddhism can say about free will (Adam 2010, Federman 2010, Frequinon 2009, Garfield forthcoming, Gier and Kjellberg 2004, Gómez 1975, Goodman 2009, Griffiths 1982, Harvey 2007, Hyland 2014, Kalupahana 1976, 1992, 1995, Meyers 2010, Rāhula 1974, Siderits 1987, 2008, Story 1976, Tuske 2013, Wallace 2008). However, and as Western philosophers of mind have come to see the nature of mind, agency, and action in increasingly impersonal terms, they have turned an eye toward Buddhism to mine its rich history of philosophical articulation of these ideas (Author 2010a, 2010b, 2010c, 2012a, 2012b, 2012c, 2014, Blackmore 2014, Coseru 2011, Flanagan 2002, Author 2010a, 2010b, 2010c, 2012a, 2012b, 2012c, 2014, and Strawson 1986, 2005). Toward this end, I defend the idea that Buddhist thought can ground a moral-responsibility-entailing kind of free agency against all four of the major skeptical arguments to be reviewed here, but let us first sketch the *autonomist possibilism* that combines all forms of autonomism and thus represents the combined target of these major skeptical arguments.

Classical compatibilists, like Hume, argued that free will is compatible with determinism because it is not the fact *that* our actions are caused that matters for morally responsible autonomous agency, but rather *how* they are caused. Classical compatibilists offer divergent accounts of how our actions are construed to be appropriately caused, but typically these accounts are articulated in *negative* terms – that is, they are typically characterized as involving the absence of freedom-undermining conditions. For example, if they are not coerced, constrained, compulsive, the result of brainwashing, torture, or manipulation, all of which are freedom-undermining conditions, then the mere fact that they are caused need not undermine the

claim that they may be sufficiently free for purposes of holding the agent who performs them responsible. *Neo-compatibilists* attempt more explicitly naturalistic, sophisticated, *positive* accounts of such matters, focusing generally on whether or not the agent exhibits the appropriate sort of *proximal control* over her action-producing processes; for example, what matters for these accounts is whether or not they are voluntary, reason-responsive, in accord with preferences or values, the result of deliberation, and so forth. Some neo-compatibilist accounts combine the absence of negative conditions with the presence of positive conditions.

Let us begin to discuss how Buddhism figures in here. Though much in Buddhism and in thinking about free will exceeds it, I focus here only on what fits within a naturalistic framework, simply so as to occupy the more parsimonious position. Posing an immediate problem for the notion of an autonomous agent, Buddhism denies the very reality of the agent or self on three related grounds – *impermanence*, *mereological reductionism* or *mereological eliminativism* (the partial or complete denial, respectively, of the reality of partite wholes), and *dependent origination*; I will explain each ground shortly. These three ideas each implies the unreality of selves and wills (free or otherwise), but also such everyday items as hands. There is also a fourth ground involving the Buddhist concept of *anātman* or non-self, that is, the Buddhist denial of the non-Buddhist Indian philosophical conception of *ātman*, the self as an unchanging, immaterial soul, but while this is a Buddhist doctrine, if not the core metaphysical insight of Buddhism, it is essentially the rejection of a non-Buddhist doctrine.

First, impermanence characterizes all conditioned phenomena: nothing in form (nothing spatiotemporal or even just temporal) lasts unchanged for more than an instant. Thus, if there is

any candidate for a self-like element, it is at best a momentary flash in a temporally-contiguous psychophysical series of shifting impersonal elements.

Second, a whole is at best just a way of designating a configuration of parts, but does not exist independently of the parts (how could it?), nor is it identical to them (for they are many and it is one): it is merely a conceptual imputation. Thus, "the self" is just a convenient way to designate the psychophysical series; pronouns and proper names basically serve the same function. Some Buddhists are reductionists about the self and others are eliminativists about it, but all Buddhists reject the ultimate reality of the self.

Third, according to the doctrine of dependent origination, which at least resembles and at most is identical with determinism, because all conditioned phenomena originate in dependence on other conditioned phenomena, no conditioned phenomenon can exist autonomously or independently. An organism, for example, cannot exist independently of all the extra-organismic, eco-system-dependent nutrients that its cellular structure requires for its functioning. Since no conditioned phenomena or beings exist autonomously or independently, all conditioned phenomena and beings are metaphysically insubstantial or *empty*, lacking any independent ontological essence, status, or *intrinsic nature*. Western philosophers might interpret this as a Buddhist claim to the effect that the self is not a *natural kind*, and, more generally, that there are no natural kinds whatsoever. (Natural kinds are sets or types of things that putatively have mind-independent features in common, such as chemical elements, as opposed to sets or types of things that only have mind-dependent characteristics in common, such as the characteristic of being called by a certain nickname.) The self, in any event, lacks any ontologically independent or substantive existence on all three grounds.

And fourth, according to the Buddhist doctrine of *anātman* or no-self, in the same way that Western philosophers have objected that the Cartesian mind, being immaterial, lacks any possible way of interacting with or influencing any physical brain or body, calling into question the coherence of the very notion of an immaterial mind itself, so too the Buddha objected that the unchanging nonphysical *ātman* would be incapable of doing anything, much less capable of accounting for and explaining volitional behavior, so there is no possible evidence for it to be found in our experience or in reason (Federman 2010). These arguments mutually reinforce each other, for whatever might putatively constitute the self cannot be causally inert, and therefore it cannot be independently existent, it must be dependently originated and thus momentary, partite, and empty. Anything partite must also be momentary and empty, and so on.

Given the Buddhist denial of the ultimate reality of self, it seems intuitive that Buddhism would equally deny the reality of autonomous agency. For it is unclear how a non-agent or non-self can be autonomous – self-governing. Charles Goodman, for instance, thinks that because there is no self, there cannot be an *autonomous self*, so there cannot be *autonomy* (2009). True, if selves do not exist, there cannot be *autonomous selves*, but that does not mean there cannot be autonomy – any more than the alleged nonexistence of red apples means there cannot be anything red (Author 2012b). Those who deny partite wholes like chariots – to use the classic Buddhist analogy – or selves may be pressed to deny the parts of parts, and so on, all the way down (Siderits 2008). But this is problematic because whatever has magnitude is logically divisible *ad infinitum*. Thus, nothing partite is reducible to anything non-partite, but *non*-partite indivisibles – if any – lack magnitude and thus cannot aggregate to form anything that could play the role of even an illusory whole. Since the world of experience is populated by such aggregated

or seemingly whole entities, this line of reasoning alone cannot account for it (Author 2012a). In other words, if this is an argument against the reality of the self, it is also an argument against the reality of chemical bonds, hands, chairs, Buddhas, and just about everything else.

Thus, attempts to rest negative conclusions about autonomy on the rejection of wholes substitute one mystery for another, and beg the question, because the premise is more problematic than the conclusion. A similar analysis may be applied, *mutatis mutandis*, to the other Buddhism-intrinsic grounds for the rejection of self, namely, dependent origination or impermanence. (The rejection of the Buddhism-extrinsic *ātman* stands, but few contemporary Western philosophers, if any, would embrace the rejected Indian/Cartesian conception.) The fact that plants cannot exist independently of the photons that fuel the photosynthesis they need does not entail that there are no plants, but only that there are no ontologically autonomous plants – plants that exist independently of their assimilation of plant-exogenous materials. The putative fact that all conditioned phenomena are at some ultimate level impermanent does not imply that there is no such thing, say, as the Verrazano Bridge, but only that its micro-constituents are in a constant state of change. I could care less if the money in my account is constantly vanishing, as long as it is also constantly being replaced almost instantaneously. Similar reasoning applies to the suffering of sentient beings, the elimination of which is one of Buddhism's prime directives.

Buddhism avoids this line of objection by dividing reality into *two truths*, *conventional* and *ultimate*. Buddhism acknowledges the conventional existence of plants, money, sentient beings, and so on, but denies that such entities exist in ultimate reality. Arguably, the *two truths* distinction maps loosely onto the difference between mereological reductionism and mereological eliminativism: conventional reality, on this view, involves what may be described

as a form of mereological reductionism about wholes, in that wholes sort of exist, somewhat, pragmatically, but only because they reduce ultimately to things that really do exist at the micro-level, and ultimate reality, on this view, involves what may be described as a form of mereological eliminativism about wholes, in that they don't really exist at all. Joining these two views or truths, we may say that there exist such everyday items as bridges, but ultimately nothing exists independently of our conceptual designation of the dependently-originated, momentary, ontologically insubstantial or empty conditioned phenomena the spatiotemporal aggregation of which we pragmatically identify as bridges and such. Thus, it makes conventional, pragmatic sense to divide the world up into chariots and agents who drive them, although ultimately these decompose into series of empty ephemera. Karma may be accounted for on this view simply as cause and effect, and reincarnation may be accounted for as an extension of the already-persisting illusion of personal survival that attends any two consecutive moments of ordinary consciousness in the same person-series – despite there being no ultimately existent self that is the bearer of karma, who reincarnates, or who persists identically through any consecutive moments (Siderits 1987, 2008).

Despite its promising appearances, the two truths doctrine (hereafter, "2Ts") admits of divergent, complex interpretations, and seems more problematic than free will does. Thus, appeal to 2Ts alone constitutes another form of question-begging. For if 2Ts is to ground a negative view of free will, as a premise 2Ts requires greater support than the conclusion it purports to support (Author 2012a, 2012b). For 2Ts not only denies the ultimate reality of free will, but also of hands, money, and just about everything else – except, on one reading (accepted only by pre-Mahayana Buddhists), the ultimate reality of *dharmas* (Sanskrit; Pāli: *dhammas*): impartite

atomistic psychophysical tropes like whiteness, sweetness, volition, and the like (Siderits 2003), which entities are, from a Western philosophical and scientific perspective, of a dubious nature (Author 2012a).

At the same time, 2Ts is analogous to the scientific bifurcation of such pairs of folk/scientific concepts as *water* and $H^2 0$, *weight* and *mass*, and *heat* and *mean molecular kinetic energy*. If all that is meant by "there is no such thing as heat in ultimate reality" is that "there is only mean molecular kinetic energy," then the analogous claim for autonomy as some complex form of volitional regulation of volitions, for example, seems hardly eliminative or unpalatable. I see no bar in Buddhism to the latter, reductionist interpretation, which I have developed in detail elsewhere (Author 2010a, 2010b), and which I sketch below.

Challenges to Free Will from Causation

The four main skeptical arguments discussed above may be grouped together with a variety of related skeptical challenges to free will as *challenges from causation*. These arguments are challenges from causation to the extent that they all involve some form of the claim that because putatively free volitional behavior is causally conditioned or influenced in some way or another it is not or cannot be free.

The consequence argument may be seen to count as the most straightforward challenge from causation. This is because it asserts that if putatively free choices and actions are caused by prior states of the universe in accordance with deterministic laws, they cannot be free because it is never open to agents to choose or do otherwise than whatever was determined by the earlier state of the universe for them to choose or do. The manipulation argument counts as a challenge

from causation. This is because it asserts that there is no difference between the case of the manipulation of the agent's beliefs, desires, or will by anything extrinsic to the agent, which obviously renders the agent's putatively free choices and actions unfree, on the one hand, and determinism, on the other hand. The randomness argument counts as a challenge from causation. This is because it asserts that even if the will is indeterministically caused it is not free because if choices are indeterministic, then the causal reigns, so to speak, that govern which choices occur are beyond the agent's control. And the impossibility argument counts as a challenge from causation. This is because it asserts that the agent's current and immediately preceding mental states condition the agent's choices, in which case they are never entirely free, but the conditioning of choices by occurrent mental states involves a kind of causal influence, however unspecified.

In addition to these four major forms of challenges from causation, there are other skeptical challenges from causation. *Illusionist* challenges from social science and neuroscience research that cast doubt on the causal efficacy of various cognitive abilities associated with the conscious mind, for example, may also be considered forms of the challenge from causation, because they claim that the conscious mind or some part of it does not play the right sort of role in the causation of putatively free choices and/or actions, but only appears to do so. For example, research supports the claim that unconscious biases explain certain choices we think were made fully freely and consciously, such as which otherwise equal supermarket items we select based on their relative placement on shelves; research also supports that when our brains are stimulated in such a way that we will reliably make a certain type of choice, and we will confabulate or reconstruct ad hoc post-facto rationalizations that allege conscious reasons of appropriate sorts

were what led to our decisions. These studies directly challenge the claim that certain of our cognitive abilities connected with putatively free choices are causally effective, and so many other, more general studies along similar lines point to so many ways in which other of our cognitive abilities are infected by illusions, all of which together indirectly undermine the idea that it is our conscious deliberative processes that cause our choices and actions. Illusionism may be grouped under impossibilism insofar as illusionist arguments all share the impossibilist idea that the way our mental states condition our choices undermines the idea that that are choices are free, and although some illusionist arguments might bypass our mental states altogether, they nonetheless imply that our mental states are causally impotent in the production of our choices.

A few words are in order about how we will proceed from here. First, I'll discuss elements of those major challenges from causation that have occupied the minds of contemporary Western philosophers, highlighting certain key points that will help render clear my subsequent rebuttals. Then I will state what I think are cognate responses to these challenges that are available from within a modified Buddhist framework – modified, that is, because I have already set aside certain Buddhist interpretations and defended others.

Optimist's and Pessimist's Dilemmas

We already have many category terms before us, possibly complicating matters, such as possibilists, soft compatibilists, and the like, but I need to introduce a few more. Some of these terms may be redundant, considered from a certain perspective, but in different contexts it pays to focus on different features of the same concepts and positions, and to be able to group positions in different ways as well. Thus, let's call believers in either the reality or the possibility

of autonomy *optimists*, and non-believers *pessimists*. Pessimists pose a certain dilemma for optimists, such that whether determinism or indeterminism is true, there can be no autonomy. Before laying out the dilemma, let us first articulate some of its essential conceptual elements.

Thus, the basic idea in the consequence argument is that if A's cause B's *deterministically*, they do so *necessarily* in accordance with exceptionless generalizations or laws; so, if determinism is true, then all behavior occurs as a necessary result of previously determined phenomena, and the future only admits of exactly one temporal event sequence that rules out all possible alternatives, and thus rules out free will. (I'll discuss it in greater detail below, but for present purposes suffice it to say that the manipulation argument counts as a second argument for the same conclusion as the consequence argument, insofar as the manipulation argument states in effect that whatever criterion that is supposed to be compatible with determinism is functionally equivalent to determinism in the sense adumbrated in the consequence argument, and thus no different, all things considered.) And the basic idea in the randomness argument is that if A's cause B's *indeterministically*, they do so *probabilistically*, in accordance with statistical generalizations that admit of exceptions, in which case it is a mere accident which particular future will emerge from a prior state of the universe, akin to the outcome of a coin toss.

With these concepts in place, consider the dilemma faced by the optimist – the *optimist's dilemma* – as a conjunction of the consequence argument's and/or the manipulation argument's deterministic challenge and the argument from randomness:

1. Either determinism or indeterminism is true.
2. If determinism is true, then agents *cannot* bring about anything not already determined (regardless of what criteria they satisfy), and thus they lack autonomy.
3. If indeterminism is true, then choices are utterly random, so agents *cannot* claim to author them, and thus they lack autonomy.

4. Thus, either way, there cannot be autonomy.

This argument may be seen, therefore, as summing up the consequence argument, the manipulation argument, and the randomness argument. It is the basis for hard incompatibilism, the claim that free will is incompatible with determinism and with indeterminism.

However, as I have argued elsewhere (Author 2010b), an opposite *pessimist's dilemma* may be posed for the pessimist by retaining claim 1, and negating and transposing the consequents across claims 2 and 3, yielding the opposite of claim 4:

1. Either determinism or indeterminism is true.
2. If determinism is true, then choices are not utterly random, so agents *can* claim to author them, and thus they may possess autonomy.
3. If indeterminism is true, then agents *can* bring about things not already determined (and satisfy criteria that make a difference), and thus they may possess autonomy.
4. Thus, either way, there can be autonomy.

The pessimist's dilemma constitutes in outline the main argument for what I refer to as *soft compatibilism*, the opposite of hard incompatibilism: soft compatibilism is the view that free will is compatible with determinism and with indeterminism.

In support of the two middle premises in the pessimist's dilemma, I need only refer to the works of other philosophers. As for the indeterminism premise, Kane, Mele, and others have given good reasons to think indeterminism may be situated in parts of the deliberative process that would *enhance* autonomy; for example, Kane constructs a model according to which conflicting deterministic processes generate an indeterministic space that opens up alternate possibilities and enables the agent to be the efficient cause of which possibility is actualized, and Mele has argued that randomness in the generation of reasons for consideration, precursors to reasons for action, would increase the creativity and diversity of such options, which intuitively would enhance the range of possible choices opened to the agent. As for the determinism

premise, Frankfurt, Dennett, Flanagan, Fischer, and others have presented plausible accounts of ways determined agents may act with varying degrees of metavolitional/volitional accord, proximal self-regulation, voluntariness, and/or reason-responsiveness sufficient for responsible agency (Kane 2002, Mele 1995, Frankfurt 1969, 1971, Dennett 1984, Flanagan 2002, and Fischer 2006).

Dispositions, Counterfactuals, Possibilities, and Actualism

Let us now turn to more directly challenge some hitherto unnoticed weaknesses in hard determinism. Arguably, just as salt is able to dissolve in liquid, an agent is able to raise his hand voluntarily. As noted above, a particularly powerful, prevalent, and seminal argument for *hard determinism* – the general claim that determinism is true and thus there cannot be free will – may be expressed in the following, shortened version of the *consequence argument*:

1. Whatever is not determined can never arise.
2. Thus, an agent can never do otherwise – anything other than what was determined.
3. Because the agent can never do otherwise, the agent cannot be autonomous.

Such a hard determinist may object to the solubility/autonomy analogy by arguing that salt cannot bring it about that it is placed in water, so it is not *able* at all. However, dispositions like solubility entail *counterfactuals* (hypothetical statements contrary to the facts) of the form, "*had* this salt been submerged in water, it *would have* dissolved," or, fairly equivalently, subjunctives of the form, "if this salt *were* submerged in liquid, then it *would* dissolve." This reduces to the claim that it is *possible* salt dissolves, if certain conditions arise, even if salt is never exposed to those conditions, and never *actually* dissolves.

A hard determinism that denies non-actual possibilities may be called *actualism*, the view that only what is, was, or will be *actual* is possible. Actualism rules out non-actual possibilities, so if autonomy requires them (say, by requiring that the agent could have done otherwise had she wanted to), then autonomy does not exist.

However, deterministic actualism cannot reject counterfactual possibilities without defeating itself, because determinism is no more than an abstraction from all counterfactual-supporting generalizations or laws; for example, A's necessarily cause B's, which entails the counterfactual claim that if there had been an A, then there would have been a B, or the subjunctive claim that if there were an A, then there would be a B. *Deterministic actualism* is thus oxymoronic; by default, *deterministic counterfactualism* is tautologous. Thus, the following determinism-friendly subjunctive (or an equivalent counterfactual) may be held to be partly constitutive of autonomy:

> If in doing x an agent *would* have done *likewise* even if he *could* have done *otherwise*, and if he *would* have done *otherwise* had he *wanted* to, then he is autonomous in doing x

Let us call this *the counterfactual principle of autonomy* or *CPA*. CPA-type principles may be used to identify a determinism-friendly model of responsible agency. For CPA may be satisfied by agents in deterministic worlds (Author 2010b).

Unlike salt, which is helpless regarding whether it ever meets its own solubility condition (of liquid immersion), agents can bring about the conditions needed to increase what Buddhist theory identifies as *mental freedom* and thus they can bring it about that they cultivate abilities that increasingly satisfy the conditions specified in CPA. And Buddhism prescribes exactly how we may do so in its many prescriptions for approaching *enlightenment* (mental freedom): by cultivating, through meditative disciplines, such things as true beliefs, wholesome volitions,

mindfulness, mental focus, mental stability, detachment, phenomenological insight into the triggers of our own volitional processes, mental equilibrium, and so on, and by decreasing such things as false beliefs, unwholesome (ego-based) volitions, mindlessness, mental disturbances (anger, jealousy, greed, hatred, craving), attachment, and so on. To the extent an agent follows the prescription he will attain mental freedom, which admits of degrees, and – this is the main point of this paper, and I will argue for it in greater detail as we proceed – to the extent that a being increasingly approximates the Buddhist state of mental freedom she increasingly satisfies CPA in her choices, thereby exhibiting increasingly effective agency or free will

Buddhist versions of CPA might be formulated to express such ideas as that if an agent, say, A, were thoroughly mindful of A's volitions, then A would be less driven by those volitions or, if A had made an effort to bring about an antidote volition V2 (say, compassion-involving) to counteract an unwholesome volition V1 (say, resentment-involving), then A would have displaced V1's volitional impetus. The less A is driven by unwholesome volitions, the more control A has over them. This increase in mental freedom is coextensive with an increase in responsible agency.

Even if responsible agency is not afforded intrinsic value in Buddhism (because Buddhism seeks to move Buddhists away from such reactive attitudes as blame, as well as away from the very idea of the agent as an autonomous controller of actions), these considerations show that the sort of responsible agency adumbrated here clearly has instrumental *soteriological* (enlightenment-oriented) value, and thus, even though Buddhism might have generic reasons to avoid abstract philosophical discussions about free will, this particular discussion of free will is not ruled out by Buddhism's general restriction of valid topics to what is of soteriological value.

Strawson's Impossibility Argument

As noted earlier, even a very opaquely described *influence* counts as a challenge from causation. Although Galen Strawson is a kind of compatibilist (Author 2012c), his impossibility argument rests on an opaquely described *influence*, one more general than the challenge reflected in the optimist's dilemma. The impossibility argument claims that the mental state one is in at the moment of choice *influences* or *conditions* that choice, rendering the choice a *function* of that mental state and thus less than ultimately free in the responsibility-entailing sense.

Strawson argues that the only way one could be uninfluenced and thus ultimately responsible for one's choice is if one began in an uninfluenced state, but an uninfluenced state would only be possible for a being that created itself *ex nihilo* or from no conditions at all, a kind of *causa sui* or self-created being, which beings are logically impossible (because nothing can be created *ex nihilo*, in general, but more specifically, in order to create oneself one must exist prior to one's creation, which is logically absurd). Strawson's argument glosses the other causal challenges, for whether one's mental state is deterministically, indeterministically, or unconsciously/neurally generated is irrelevant to the fact that it will influence one's choice. I take it therefore that his impossibilist argument subsumes all the other skeptical, pessimistic arguments discussed above, the illusionist's included, since whether one's choice is influenced in a deterministic, indeterministic, or illusionist manner is irrelevant to the fact that it is influenced or conditioned. Thus, any rebuttal of the impossibility argument is thereby also an argument for possibilism, and may double as a rebuttal to all the other skeptical, pessimist arguments and positions. I adduce such a Buddhist rebuttal to this inclusively-construed impossibility argument

shortly. Again, though Buddhism has historically ignored the free will issue and some contemporary Buddhist scholars prefer to continue to do so, the mere fact that Buddhism has the conceptual tools to be able to formulate a possibility argument that can rebut all these challenges from causation is reason alone not only to consider it, but to take it quite seriously.

Frankfurt's Distinction, Meditation, and Freedom-fostering Influences

In my view, a Buddhist conception of free will may be formulated negatively, as a kind of freedom *from*, rather than positively, as a freedom *to* – that is, as a freedom from various forms of mental bondage, particularly delusional cognitive and volitional bondage. These are related in obvious ways when, say, we are mindlessly attached to satisfying unwholesome desires based on cognitive errors. According to Buddhist theory, all ego-based volitions exemplify this relationship between ignorance about the nature of the self and its corollary volitional bondage.

Practices that increase such bondage generate *mental-freedom-undermining influences*; practices that decrease it generate *mental-freedom-enhancing influences*. Buddhism posits that meditation reduces mental-freedom-undermining influences, and thus increases mental freedom, precisely through mental state *influences*. These mental state influences may be viewed usefully through the lens of Frankfurt's distinction (1971) between first-order desires (desires for experiences or things) and higher-order desires or metavolitions (desires that have other volitions or desires as their objects, such as a desire that certain volitions are effective and that others are not, or pro- and con-attitudes towards certain first-order desires).

The *will* of a person is typically volitionally/metavolitionally structured or hierarchical, according to Frankfurt, whereas the will of a nonhuman animal or infant human is typically

unstructured, entirely first-order. *Weakness of will* afflicts only hierarchically structured wills, when metavolitions fail to regulate volitions in accordance with the agent's metavolitional preferences. Chronic cases of weakness of will result in addiction, compulsion, and other breakdowns of will. Despite its lack of adherents, and certain technical objections (Tuske 2013), these elements of Frankfurt's model are insightful, explanatorily powerful, and highly intuitive (Author 2010b). To illustrate this latter claim, elements of Frankfurt's model shall be shown to fit nicely with a certain Buddhist conception of freedom. Let me emphasize that I am not arguing that Buddhism endorses a Frankfurt-style theory of free will, but only that it may adapt Frankfurt's distinction to fruitful ends.

Thus, Buddhist meditators are increasingly able to form the sort of mental-freedom-oriented wills they want, through long, disciplined effort. Buddhist meditative discipline proceeds essentially along two related lines of long-term practice: *one-pointedness* (concentration) and *mindfulness* (paying attention to, or being mindful of, what one is focusing on). These terms have technical meanings in Buddhism that are sometimes conflated in English. *Mindfulness* is the translation for *sati* (Pāli; Sanskrit: *smṛti*); mindfulness is thought to lead to insight. *Insight* is the translation for *vipassanā* (Pāli; Sanskrit: *vipaśyanā*), although popular Anglophone usage of *vipassanā* loosely translates it as *mindfulness*. *One-pointedness* is the translation for *samādhi*, which is thought to lead to tranquility, calm abiding, or mental quiescence (*śamatha*). Meditative *discipline* is also the translation for *samādhi*, which includes mindfulness and one-pointedness as well as effort. (Lutz, Dunne, and Davidson 2007.)

This discipline is the introspective analogue of intense athletic (proprioceptive) training of the body or of scientific (exteroceptive) training of analytic observation. Buddhist meditation

cultivates a heightened degree of impartial, non-judgmental (but highly discriminative), non-reactive, concentrated focus on and attention to the phenomenology of the practitioner's cognitive/conative dynamics, which practice renders those dynamics more transparent to the agent. This long-term phenomenological investigation generates detachment from spontaneously generated volitional impulses – that is, an ability to not act on the sorts of first-order volitions that typically lead to action.

Skill at introspective investigation of volitional phenomenology generates insight into the impersonal nature and functioning of volitional phenomena, and thereby fosters the meditator's ability to dis-identify with egocentric volitions, to not act on them, and to increase control over them, as well as the meta-ability to form appropriate, effective pro- and con-attitudes or metavolitions regarding freedom- and bondage-fostering volitional influences, respectively. That is, by cultivating heightened levels of introspective skill, long-term meditation practitioners become increasingly mindful of the functional dynamics of their volitional dispositions, increasingly able to refrain from acting on volitional impetuses, increasingly practiced at fostering those cognitive/conative influences that are freedom-enhancing, and diminishing those that are bondage-enhancing. All of this is *prima facie* consistent with the 2Ts conception of the conventional agent (the volition-regulating, hierarchically complex meditation practitioner) as an ultimately impersonal series of psychophysical phenomena, technical difficulties associated with the 2Ts and with Frankfurt's two-tiered model notwithstanding.

The problem with the move up to a higher level in a volitional hierarchy, say, L2, has been claimed to imply that the lower-level volition, L1, is free if and only if L1 is endorsed at L2, but the same requirement is then pressed on L2, generating a move to L3, and so on, such that

this iterative formula generates an infinite regress. Even if this is fatal to Frankfurt's model, it is not fatal to my counterfactual version of it (Author 2010b). Because the Buddhist model does not need an iterative formula, but only the claim that metavolitional CPAs correlate increases in mental freedom with increases in responsible agency, it avoids Frankfurt's regress problem (Author 2010a).

The Fallacy of the Heap

The Frankfurt-informed Buddhist model of responsible agency formulated above may be used to undermine Strawson's impossibility argument. The fallacy of the heap is a sorites type fallacy because it erroneously infers that there are no such things as heaps of sand from the inability to demarcate an exact number of grains of sand within a range of *n+1/n-1* grain that would be required to constitute or identify a heap or non-heap of sand, respectively.

Just as there are obviously heaps of sand and non-heaps, though we may not be able to demarcate them by any exact value of *n* for the formula of *n+1* or *n-1* grain, there are obviously more or less free minds and choices, though they emerge from within conditioned or *influenced* mental states and we have yet to demarcate their exact threshold conditions. But Buddhism comes close in the *Abhidharma*, Buddhaghosa's *Path of Purification*, and other texts (Bodhi 1995, Buddhaghosa and Nanamoli 2003), for example, which texts specify the stages on the path to *nirvāṇa* (Sanskrit; Pāli: *nibbāna*), or enlightenment.

Skilled in introspective phenomenological investigation into the dynamics of volition, long-term Buddhist meditation practitioners are increasingly able to refrain from acting on the sorts of mental-bondage-fostering volitional impulses that typically prompt unskilled, ordinary

individuals into action, further diminishing their mental (cognitive/volitional) freedom, if not to fully detach from those impulses and eliminate them in favor of antidote-type mental-freedom-fostering volitions. Surely, this skill places such practitioners toward the higher end on the scale of mental-freedom-fostering and mental-bondage-fostering *influences*. We need not be able to quantify the extent of their mental freedom in precise terms, just as we may not be able to numerically define the number of grains of sand required for sand heaps, but surely there are small, medium, and large heaps of sand as well as persons with lesser and greater degrees of mental freedom.

The impossibility argument seems to demand that *any* degree of mental-state influence entails mental bondage or unfreedom, but this seems wildly implausible in itself, and much more so in light of the above considerations. Even if we *grant* the point, however, which we need not, Buddhism offers the possibility of attaining total mental freedom upon the attainment of enlightenment or *nirvāṇa*, which may be defined for our purposes here as liberation from all ego-volitional impulse and cognitive delusion – *all mental-freedom-undermining mental state influences*.

Since Buddhism posits that unenlightened beings – beings caught within the realm of mental bondage – may become enlightened, particularly through cultivation of the sort of meditation-based volitional/metavolitional self-regulation discussed above, Buddhism implicitly rejects the impossibility argument: mentally unfree beings are beings whose choices and actions are influenced in mental-bondage-fostering ways by certain mental states, but they can attain mental states that reverse and extend beyond the reach of such influences and that completely extinguish them. This is clearly *not* thought to require any *ex nihilo causa sui*. Thus, the Buddhist

model is tantamount to a model of how what the impossibility argument says is impossible is actually possible. It also models how that argument commits a sorites (or otherwise equivalent) fallacy.

Dharma-responsive Responsible Agency

The Buddhist teachings, path, or way may be referred to as the *Dharma*, which term more generally denotes the soteriologically relevant dimension of the truth (according to Buddhist understanding). If an agent is sufficiently sensitive and responsive to the truths of impermanence, insubstantiality, and impersonality as well as to the suffering of sentient beings, then she is what I have described as *Dharma-responsive* (Author 2010a). *Dharma*-responsiveness is a species of what compatibilist and *semi-compatibilist* philosophers like Wolf (1990) and Fischer (2006), respectively, have described as *reason-responsiveness*, and have identified as playing a central role in determining whether and to what extent an agent exhibits any determinism-friendly responsible agency.

Semi-compatibilists separate moral responsibility from free will: they are hard determinists about free will because they think it requires indeterminism, but soft determinists about moral responsibility because they think reason-responsiveness suffices for it and is compatible with determinism. Reason-responsiveness may be easily understood by analogy with wind-responsiveness: a weathervane that functions according to its intended purpose is sensitive to the flow of wind as well as responsive to it, whereas a malfunctioning weathervane (for example, a rusted one) is neither. Likewise, a rational being recognizes and is sensitive to relevant reasons in varying contexts and responds to them appropriately in choice and action.

Wolf, Fischer and other compatibilists and semi-compatibilists argue along slightly different lines to the shared conclusion that even if agents lack the kind of autonomous ability to do otherwise under identical conditions that is impossible in a deterministic world (but which might be possible in an indeterministic world or under a dualistic model), to the extent that they are moderately reason-responsive, they have enough of a compatibilist kind of free will that renders them morally responsible for their actions, precisely because they are functional moral agents, that is, agents that are able to recognize and respond appropriately to moral reasons.

Insofar as *dharmic* (*Dharma*-based) reasons are a species of reasons, *Dharma*-responsiveness is a species of reason-responsiveness. Thus, to the extent that reason-responsiveness grounds attributions of compatibilist responsible agency, so does *Dharma*-responsiveness. This is a logical point, but there are also substantive features of *Dharma*-responsiveness that establish the same conclusion on extra-logical, Buddhist-intrinsic grounds. For a *Dharma*-responsive agent is independently sensitive to a complex set of Buddhist moral reasons that concern whether entities will suffer and excludes their biological and related attributes (species, gender, and so on) or whether they are self-related, rendering them genuinely altruistic reasons. Thus, such a *Dharma*-responsive agent is not only a responsible agent, but robustly so. Responsible agency here may be construed as genuinely metaphysical, not merely utilitarian in the sense critics intend by the idea that fostering attitudes, norms, and institutions of *holding* people responsible – even though, metaphysically, they are not responsible, and even by Skinnerian means – has social utility. For this sort of responsible agency involves a complex array of CPAs the satisfaction of which indicates that the agent instantiates volitional regulation and thus possesses a degree of what may be considered Buddhist free will.

From the soteriologically framed perspective of Buddhism, the more general abilities that constitute reason-responsiveness only matter insofar as they support *Dharma*-responsiveness, though the same soteriological instrumentality limits the importance of this Buddhist form of free will. Nonetheless, despite its intrinsic unimportance, this form of free will is instrumentally essential to the entire Buddhist soteriological project, despite also the fact that upon attainment of the ultimate Buddhist goal of *nirvāṇa*, what constitutes the agent – in an important sense, namely, ego-volition-functionality – ceases to exist. But this is the attainment of total freedom, paradoxically: a kind of agentless agency.

Autonomous Non-agents? The Deflationary Self

This model of responsible agency is consistent with deflationary models of the self, as articulated by Siderits, Strawson, Flanagan, and others: the self is not a *thing*, but a densely clustered set of causal *processes* that are significantly reflexively looped within the cognitive/conative, metacognitive/metavolitional dimensions of the ultimately impersonal psychophysical stream that is a person-series and that significantly shapes itself. The meditator who cultivates effective metavolitions is highly responsible for shaping and sculpting the sort of self-process she becomes over time, and the more she does, the greater her responsible agency, though the more she advances the less substantive, static, or real her *self* or her autonomy seems *to her*.

Theoretically, maximal mental freedom thoroughly eliminates the illusion of the separateness of the self-process associated with responsible agency. But in total mental freedom there's no need for *self*-regulation, as there is no longer any separate, ego-volition-constituted

psychological complex driving the person-series. Enlightened action is spontaneously dharmic, in harmony with the *Dharma*, the way ultimate reality is.

Socratic Buddhist Wisdom: Awareness of Ignorance and Illusion

Buddhism, philosophy, and science share the aim of uncovering ways in which our mental lives are subject to illusion, in favor of a more ultimate reality. Buddhism *admits* we are mostly unfree because we are, more frequently than not, in the grip of mental states that are deluded about the self and related misconceptions (permanence, substantiality, wholeness, and so on) and thus filled with imbalances of volitional attraction or aversion such as greed and hatred, respectively.

From the ultimate or enlightened perspective, everything conditioned, relative and conventional is relatively illusory, but that does not eliminate important distinctions within the relative domain, such as that between freedom and bondage, wholes and parts, or those who have hands and amputees. Rather, the sort of relative, proximal autonomy Buddhism cultivates is necessary to attain the non-illusory perspective of total mental freedom, enlightenment.

The Consequence Argument, Neo-compatibilism, and the Manipulation Argument

This Buddhist conception of freedom fares well against the impossibility argument, as we have seen, but also against some of the other, very powerful forms of autonomy pessimism, such as the consequence and manipulation arguments. We have seen a shortened version of the consequence argument already, but let us recall it, to segue into the more challenging manipulation argument, which builds on the consequence argument idea that determined behavior cannot be free.

The consequence argument reduces to the claim that determinism leaves open only a single event-series or single outcome in each moment, thus an agent can never bring about anything other than what is already a determined consequence of past conditions together with the laws of nature, so he cannot do other than he does, he lacks autonomy, and he cannot be morally responsible (van Inwagen 1975). But neo-compatibilists like Frankfurt, Dennett, and Flanagan argue plausibly – albeit along slightly divergent lines – that causation is consistent with some version of responsible agency, along lines loosely similar to those just sketched about metavolitional/volitional regulation, reason-responsiveness, and so forth. This is likely why incompatibilists tend to rely more heavily, as many, like Pereboom, now do, on the manipulation argument.

The manipulation argument asserts, most generally, that whatever criteria the compatibilist produces as sufficient for autonomy can be indirectly designed so as to be satisfied by the agent *proximately* but *ultimately* by a hidden manipulator, and that determinism is functionally analogous to, or morally indistinct from, such manipulation. In the classic version, Pereboom (2001) adduces a four-case version of the argument: in the first case, say, case A, a neuroscientist remotely manipulates an agent's neurons, rendering the agent unfree (even if the agent apparently satisfies various compatibilist criteria); in case B, the agent's brain was modified and programmed during infancy in such a way that it would lead to the agent's making the same choice as in case A (and satisfy the same criteria); in case C, circumstances in the agent's environment are manipulated so as to lead the agent to make that same choice; and, in case D, determinism is true. Pereboom challenges the optimist to identify a morally relevant,

principled difference between any two adjacent cases, claims there cannot be any, and thus that determinism is no different from manipulation.

Various versions of the argument identify various forms of manipulation, internal, external, intentional, and otherwise, and some versions order them in a sort of sliding scale of intuitions, but because the agent is only proximally but not ultimately responsible for the fact that she satisfies the criteria in either case (in one set of cases, via manipulation, and in the other case, via deterministic means), all of these versions of the argument place the burden on the compatibilist to identify a relevant difference between the ways in which such forms of manipulation are autonomy-undermining and the way in which determinism is not autonomy-undermining (Vihvelin 2011). This is considered by many to be the strongest argument against compatibilism. What can the Buddhist say in response?

Free-will-possessing and Enlightened Brains in Vats

Suppose the Buddhist practitioner progressing along the Buddhist path has satisfied sufficiently many Buddhism-specific CPA-type principles, whatever they may be, thereby rendering that practitioner significantly metavolitionally self-regulating, *Dharma*-responsive, and so on, thus exhibiting what I have alleged may be considered a rather robust Buddhist form of free will. Suppose further that this individual was completely manipulated by agent-extrinsic means: conjure up whichever type of manipulative scenario that suits your 'phi-fi' imagination (neural implants, cult brainwashing – whatever). (I coined the neologism *phi-fi* for *philosophical fiction*, akin to *sci-fi* (Author 2010b).) In the same way that these sorts of conditions are obviously the manipulative causes of the agent's *merely proximal* satisfaction of the putative autonomy criteria

and thus the agent only appears to be autonomous, so too if determinism is true, then determinism *necessarily* triggered a set of functionally equivalent, ultimately agent-extrinsic conditions that brought it about that the agent proximally satisfied those compatibilist autonomy conditions, in which case – by parity of reasoning – it only appears that the Buddhist is autonomous in a deterministic world, but really is no more autonomous than those directly manipulated agents in Pereboom's four-case argument.

This seems even more pressing in the face of the Buddhist denial of the ultimate reality of the self, for it is bad enough for a non-Buddhist to face a challenge to the effect that the putative agent is really just a puppet whose strings are controlled by someone else, whereas for the Buddhist there is not even a puppet to be controlled by others, but only strings controlled by impersonal forces and processes. The burden thus appears to be squarely on the Buddhist to identify a relevant difference between impersonally/deterministically caused proximal satisfaction of autonomy criteria and their manipulated proximal satisfaction.

The Buddhist model of freedom that I have been formulating, however, may easily respond to this argument, and more easily, actually, than the more generic compatibilist or soft determinist can. Suppose the neural correlates of a robust form of autonomy or even of *nirvāṇa* could be directly generated by neuroscientific manipulation. Would that undermine them? *Absolutely not!* With manipulative enemies like that, one would not need friends. In other words, whether my satisfaction of the metavolition-regulating, mental-freedom-generating, and related CPA-type Buddhist counterfactuals or my partial or total mental freedom was brought about by manipulation, determinism, indeterministic chance, the Buddha's blessings, God's grace, my meditative efforts, or anything else whatsoever does not matter at all. And that is because

whatever has led me to be free has freed me! After over 40 years of often tremendously difficult meditation practice of my own, not to mention endless struggles with unwholesome volitions and the like in unrelentingly undermining daily routines, I would be among the first to volunteer to be so manipulated.

The soft determinist or generic compatibilist may need to differentiate between manipulation and determinism, to the extent such a particular theory is committed to backward-looking or historical considerations (that is, to identify those features of its causal history that explain how it came to act freely). For the autonomous agent, on such a model, is construed to have a kind of independent identity and integrity as a relatively/partly self-constituting being, a construction that is intuitively jeopardized by direct or indirect (historical) manipulation.

The Buddhist, however, has no such commitments, is not vulnerable to such worries, and thus is under no burden to specify backward-looking principles that can differentiate between the two types of genetic cases. Likewise, libertarians – those who embrace indeterminism to avoid the consequence and manipulation arguments – are more vulnerable to a genetic objection about the chance origination of their putatively free choices (the argument from randomness), whereas Buddhists can accept a kind of fortunate enlightenment, though doctrinally they happen to reject *chance* because the Buddha rejected it.

With or without such *manipulation*, once one has become a skilled meditator, one is able to cultivate mindfulness of volitional states (and many other cognitive states) sufficient to enable one to detach from their influences, whatever the origins of those states. The skilled meditator, therefore, having reached an advanced stage in practice, whether by natural, manipulative, or

other fortuitous means, instantiates an established state of mental quiescence, detachment, and clarity sufficient to maintain equanimity and remain unperturbed and unaffected by the arising and fading of all other mental states, regardless of whether those states occur randomly, deterministically, or as a result of direct or indirect manipulation. In fact, certain Buddhists advocate the use of fear and other manipulative devices as *skillful means*, precisely on the grounds that, all things considered, doing so will alleviate mental bondage (Goodman 2009, Shantideva and Padmakara 2008). It would not matter for purposes of this argument if such a being were a cyborg fresh off an assembly line, for if once such a being is activated it instantiates the requisite causal/functional cognitive/volitional abilities, apart from being as close (functionally) to a *causa sui* as possible, that cyborg is as free as its naturally-arising cousin.

If Buddhism is right about the possibility of cultivating mental freedom, then the meditation master can escape the intensely manipulative influences even of psychic ventriloquists of the sort Descartes imagines with his demonic genius, or the physical equivalent as depicted in brain-manipulation and brain-in-a-vat scenarios. For it does not matter to such a being how an *adharmic* (unwholesome) mental state arises or by what or whose influence: what matters is whether or not the mental state ought or ought not to be approved and/or acted upon. The advanced Buddhist does not differentiate between thoughts, volitions, or other mental states that are *mine* versus *not mine* in general anyway, so the typical vulnerability of manipulation victims to thoughts, volitions, or other mental states that *appear* to be *mine* simply does not arise.

Indeed, in principle, if it is theoretically possible that a brain placed in a vat of gelatinous fluid can still have a functional mind, then it is equally possible that a skilled Buddhist

mind/brain in a vat can exhibit responsible agency and can also reach *nirvāṇa*. For such a bodiless mind whose brain is in a vat, call it *BIV*, can still exhibit *Dharma*-responsiveness insofar as it remains receptive and responsive to dharmic reasons, albeit in a virtual world. According to the standard BIV thought experiment, such a mind cannot tell the difference between being a BIV wired up to virtual senses, virtual motor nerves, virtual limbs, and virtual perceptual fields constructed by algorithms and the like, and being a brain in a skull, call it *BIS*, connected to actual physical senses, motor nerves, limbs, and perceptual environment. Nothing changes phenomenologically – *ex hypothesi* – between BIS and BIV. Indeed, according to Buddhism, being BIS is equivalent to being BIV because *the world* is functionally equivalent to a virtual reality anyway insofar as it is significantly conceptually constructed (Garfield 2011).

If BIV were fed experiences that presented BIV the *false* impression that BIV had attained that state, this might appear to undermine Buddhist confidence in this line of argument. But technically it does not matter if anyone *subjectively* but falsely *thinks* they satisfy a criterion. That is no counterexample for an externalist account such as ours. What matters is the externalist condition that the mind in question *truly* satisfies the criterion, not the (true or false) internal impression to the mind itself that it does. Thus, this variation on the thought experiment may be set aside.

What matters, then, is whether or not a mind genuinely instantiates the functional abilities in question, not a momentary mental state impression to itself that it does, nor the historical or genetic elements of how it came about that it does. This is particularly clear in the case of enlightenment, not to be confused with pseudo-enlightened momentary states of altered consciousness – something that often fools long-term meditation practitioners. The Buddha

claims that on the eve of his enlightenment he escaped the (BIV-equivalent and evil-demon-equivalent) deceptions of Mara, the archetypal Buddhist cosmic deceiver, akin to Descartes's evil demon. Only, unlike Descartes, the Buddha did not conclude that the *I* that thinks is real. Rather, he saw through it as just another deception – indeed, the greatest, whereby he attained freedom *from* the self, from its will, and from all its manipulative influences. Once freed, *nothing* can manipulate the enlightened being.

Challenge from Illusionism

On general illusionist grounds one may doubt that any mental states whatsoever, meditative or otherwise, play any truly causally efficacious role in the sub-personal neural processes leading to choices and actions, given all the evidence of error associated with so many conscious phenomena, and the fact that meditative states are altered states only seems to add to their dubious status. But illusionism goes too far if it concludes that *all* mental states are causally impotent on the grounds that *some* of them are (Balaguer 2013).

Besides, neuroscience is not only open to the possibility that the Buddhists' claims about the causal efficacy of meditative practices may be confirmed or disconfirmed, if we find the neural correlates of these mental states and abilities, but initial evidence along such lines is already promising. For example, studies performed on Zen meditation practitioners, ever training to be ceaselessly responsive to the experiential aspects of the present moment, reveal that when they are presented with repetitive stimuli their brains do not show any reduced responsiveness to the repeated stimuli, whereas the brains of those in the control group typically siphon off otherwise meaningless repetitive stimuli after a fairly short exposure period (Kasulis 1985).

Studies performed on veteran practitioners of loving-kindness meditations revealed far greater neuronal mass and activation in the left pre-frontal cortex, the so-called empathy center of the brain, than in the control group, during such meditations (Lutz, Dunne, and Davidson 2007). Studies reveal similar neuronal alterations in the brain correlated with conscious meditative practices designed to discipline attention (Ricard 2011). Finally, studies not connected at all with meditation but rather simply with contrasting cases of mindfulness and mindlessness confirm similar results (Langer 2011). While these studies are too premature to ground any confidence in the claim that the sort of model proposed here is true, they suffice to support the claim that the model is not only empirically possible, but initially plausible.

Conclusion

Although this model of free will is predicated on the limit case of mental freedom, *nirvāṇa*, in light of the initial empirical confirmation of meditative practices, it is arguably plausible that to the extent anyone approximates that limit they instantiate some modicum of free agency, whether or not their mental state at the moment of choice is deterministic, manipulated, or conditioned.

Acknowledgement

The writing of this paper was supported in part by a grant from the National Endowment for the Humanities 2012 Summer Institute, "Investigating Consciousness: Buddhist and Contemporary Philosophical Perspectives."

Citations

Abelson, Raziel. 2013. *Common sense morality*. New York: Global Scholarly Publications.

Adam, Martin. 2010. No self, no free will, no problem: Implications of the *Anattalakkhaṇa Sutta* for a perennial philosophical issue, *Journal of the International Association of Buddhist Studies* 33, 239–265.

Author. 2014.

Author. 2012c.

Author. 2012b.

Author. 2012a.

Author. 2010c.

Author. 2010b.

Author. 2010a.

Balaguer, Mark. 2014. *Free will*. Cambridge, MA: MIT Press.

Blackmore, Susan. 2014. Like science, the Buddha's ideas challenge everything we like to believe. In *Science & religion: 5 questions*, edited by Gregg Caruso, 9-23. Automatic Press.

Bodhi, Bhikkhu, ed. 1995. *A comprehensive manual of Abhidhamma: The Abhidhammattha Sangaha of Acariya Anuruddha*. Access to Insight website. Online: http://www.accesstoinsight.org/lib/authors/bodhi/abhiman.html (accessed July 24, 2014).

Buddhaghosa, Bhadantācariya and Bhikkhu Nanamoli. 2003. *Visuddhimagga: the path of purification.* Onalaska: Pariyatti Publishing.

Coseru, Christian. 2011. Practical reason and the spontaneity of compassion. Panel presentation, Contemporary Perspectives on Buddhist Ethics: Freedom and Responsibility. Columbia

University, October 6-7, 2011. Online: http://www.cbs.columbia.edu/buddhist_ethics/panel-five.html (accessed July 24, 2014).

Dennett, Daniel. 1984. *Elbow room: the varieties of free will worth wanting.* Cambridge, MA: MIT Press.

Federman, Asaf. 2010. What kind of free will did the Buddha teach? *Philosophy East and West* 60, 1-19.

Fischer, John Martin. 2006. *My way: essays on moral responsibility.* New York: Oxford University Press.

Flanagan, Owen. 2002. *The problem of the soul: two visions of mind and how to reconcile them.* New York: Basic Books.

Frankfurt, Harry. 1969. Alternate possibilities and moral responsibility, *Journal of Philosophy* 66, 829-839.

Frankfurt, Harry. 1971. Freedom of the will and the concept of the person, *Journal of Philosophy* 68, 5-20.

Friquegnon, Marie. 2009. Buddhism and free will. Columbia Society for Comparative Philosophy, April 2009. Online: http://www.cbs.columbia.edu/cscp/marie-friquegnon/ (accessed July 24, 2014).

Garfield, Jay. Forthcoming. Just another word for nothing left to lose: freedom of the will in Madhyamaka. In *Freedom of the will in a cross-cultural perspective*, edited by M. Dasti and E. Bryant. New York: Oxford University Press.

Garfield, Jay. 2011. I am a brain in a vat (or perhaps a pile of sticks by the side of the road). Draft of September 4, 2011. Online: http://www.smith.edu/philosophy/docs/garfield_brain.pdf (accessed July 24, 2014).

Gier, Nicholas F. and Paul Kjellberg. 2004. Buddhism and the freedom of the will: Pali and Mahayanist responses. In *Freedom and determinism*, edited by Joseph Keim Campbell, Michael O'Rourke, and David Shier, 277-304. Cambridge, MA: MIT Press.

Gómez, Luis O. 1975. Some aspects of the free will question in the Nikāyas, *Philosophy East and West* 25, 81-90.

Goodman, Charles. 2009. *Consequences of compassion: an interpretation & defense of Buddhist ethics*. New York: Oxford University Press.

Griffiths, Paul J. 1982. Notes toward a Buddhist critique of karmic theory, *Religious Studies* 18, 277-291.

Harvey, Peter. 2007. 'Freedom of the will' in the light of Theravāda Buddhist teachings, *Journal of Buddhist Ethics* 14, 35-98.

Hyland, Terry. 2014. Mindfulness, free will and Buddhist practice: can meditation enhance human agency? *Buddhist Studies Review* 31:1, 2014.

Kalupahana, David J. 1995. *Ethics in early Buddhism.* Honolulu, Hawaii: University Press of Hawaii.

Kalupahana, David J. 1992. *A history of Buddhist philosophy: continuities and discontinuities.* Honolulu, Hawaii: University of Hawaii Press.

Kalupahana, David J. 1976. *Buddhist philosophy: a historical analysis*. Honolulu, Hawaii: University Press of Hawaii.

Kane, Robert. 2002. Some neglected pathways in the free will labyrinth. In *The Oxford handbook of free will*, edited by Robert Kane, 406-437. New York: Oxford University Press.

Kasulis, T. P. 1985. *Zen action, Zen person*. Honolulu, Hawaii: University of Hawaii Press.

Lutz, Antoine, John D. Dunne and Richard J. Davidson. 2007. Meditation and the neuroscience of consciousness: an introduction. In *The Cambridge handbook of consciousness*, edited by Philip David Zelazo, Morris Moscovitch and Evan Thompson, 499-551. Cambridge: Cambridge University Press.

Mele, Alfred. 1995. *Autonomous agents: from self-control to autonomy.* New York: Oxford University Press.

Meyers, K. 2010. *Freedom and self-control: freedom in South Asian Buddhism.* PhD thesis, University of Chicago, Chicago, USA.

Pereboom, Derk. 2001. *Living without free will.* Cambridge: Cambridge University Press.

Rāhula, Walpola. 1974. *What the Buddha taught.* New York: Grove Press.

Ricard, Matthieu. 2011. This is your brain on mindfulness. In *The mindfulness revolution*, edited by Barry Boyce, 126-135. Boston: Shambhala.

Shantideva and Padmakara Translation Group. 2008. *The way of the Bodhisattva.* Boston, MA: Shambhala.

Siderits, Mark. 1987. Beyond compatibilism: a Buddhist approach to freedom and determinism, *American Philosophical Quarterly* 24:2, 149-159.

Siderits, Mark. 2008. Paleo-compatibilism and Buddhist reductionism, *Sophia* 47:1, 29-42.

Story, Francis. 1976. *Dimensions of Buddhist thought: essays and dialogues.* Kandy, Sri Lanka: Buddhist Publication Society.

Strawson, Galen. 1986. *Freedom and Belief*. Oxford: Clarendon Press.

Strawson, Galen. 2005. On 'freedom and resentment.' In *Free will: critical concepts in philosophy*, edited by John Martin Fischer, 58-87. London: Routledge.

Tuske, Joerg. 2013. The non-self theory and problems in the philosophy of mind. In *Companion to Buddhist philosophy*, edited by Steven Emmanuel, 419-428. Wiley-Blackwell.

Van Inwagen, Peter. 1975. The incompatibility of free will and determinism, *Philosophical Studies* 27, 185-99.

Vihvelin, Kadri. 2011. Arguments for incompatibilism. In *The Stanford encyclopedia of philosophy*, edited by Edward N. Zalta. Online: http://plato.stanford.edu/archives/spr2011/entries/incompatibilism-arguments (accessed August 7, 2012).

Wallace, B. Alan. 2008. A Buddhist view of free will: beyond determinism and indeterminism. In *Neurosciences and free will*, edited by Robert Pollack, 59-69. New York: The Earth Institute at Columbia University.

Wolf, Susan. 1990. *Freedom within reason.* New York: Oxford University Press.