Research Article

Buddhist Meditation and the Possibility of Free Will

Rick Repetti

Department of History, Philosophy & Political Science, Kingsborough Community College, 2001 Oriental Blvd., D-223, Brooklyn, NY 11235, 718-368-5226, USA.

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 culminates in actions as an ultimately impersonal, 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, independently of whether the world is deterministic. This is not to suggest that Buddhism contains or endorses a theory of free will, but that Buddhism may formulate such a theory.

Powerful arguments against the possibility of free will have become dominant in contemporary Western analytic philosophy. Buddhism has resources that may be deployed to formulate a position that may stand up against four of the most powerful arguments for free will skepticism. I support this claim in two stages. First, I address the four major arguments for free will skepticism, after outlining necessary terminology and positions shaping the debate. I argue that one such argument may subsume the others, but I address them all because the subsumption claim may be disputed, the others argue for the same conclusion, and each might survive if the subsuming argument fails. Second, I show how Buddhist ideas support a critique of those arguments. Although not entirely linear, the explication is arranged for progressively complex comprehension. Also, I focus more on these four specific Western arguments than I focus on specific Buddhist analogues because I intend to show Western philosophers that 'Buddhism'—more realistically, a variety of forms of Buddhism—has the resources for rebutting those arguments.

Comparing specific Western philosophers with 'Buddhism' nonetheless appears asymmetrical, but where appropriate I identify primary and/or secondary Bud-

dhist sources for more specific examination. The Buddhism-based account of free will formulated below is endorsed by some Buddhist philosophers and disputed by others. (Repetti (2016) contains a comprehensive collection of such positions.)

The Free Will Terrain

I present a fairly comprehensive introduction to the terminology, concepts, arguments, and issues at stake in the free will debate, but I can only say so much here. For a greater introduction to the big picture regarding free will, see Balaguer (2014), Caruso (2013), and Abelson (2013). I prefer the term *autonomy* to *free will*. *Autonomists* believe we sometimes exhibit a type of autonomy sufficient for appropriately holding us morally responsible for our actions; *autonomy skeptics* deny autonomism. *Determinists* think every event is the lawfully necessary consequence of previous events; *indeterminists* reject determinism. *Compatibilists* think autonomy is compatible with determinism. Compatibilist determinists are *soft determinists*. *Incompatibilists* think autonomy is incompatible with determinism (which excludes alternatives they consider necessary for autonomy). Incompatibilist autonomists are *libertarians*; incompatibilist determinists are *hard determinists*. *Hard indeterminists* think autonomy is incompatible with indeterminism; *soft indeterminists* think autonomy is compatible with indeterminism. *Hard incompatibilists* consider both determinism and indeterminism incompatible with autonomy. *Impossibilists* think autonomy is impossible independently of whether determinism is true, because every choice is conditioned by the mental state occurrent at the moment of choice. *Illusionists* are autonomy skeptics who think our experiential grounds for belief in autonomy are infected by subpersonal processes subject to illusion. *Possibilists* consider autonomy possible regardless of what mental state one is in; *soft compatibilists* think autonomy is compatible with determinism and indeterminism. *Autonomy skeptics*, *pessimists*, or *non-autonomists* include impossibilists, hard determinists, hard indeterminists, hard incompatibilists, and illusionists; *autonomy gnostics*, *optimists*, or *autonomists* include possibilists, soft determinists, soft indeterminists (libertarians), and soft compatibilists.

One of the most powerful, seminal incompatibilist arguments, first proposed by Ginet, is van Inwagen's *consequence argument*, which asserts that if determinism is true, agents can never do other than what they are determined to do, and thus cannot be free (Ginet 1966; van Inwagen 1975). Although van Inwagen argues only

for incompatibilism, hard determinists add the premise that determinism is true (at the non-quantum level, based on the bulk of non-quantum-level science), and conclude there is no autonomy. Libertarians are autonomists who, conversely, add the premise that we possess autonomy (for various reasons, e.g., experience, phenomenology, and its necessity for morals and other basic features of our humanistic self-conception), concluding that determinism must be false, at least of some central feature(s) of human agency, even if true of everything else.

Another powerful skeptical argument against autonomy is Pereboom's *manipulation* or *four-case argument* (Pereboom 2001). 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 advocates hard incompatibilism, which adds to the manipulation argument the hard indeterminist's *randomness argument* that if indeterminism is true, choices are random and not truly "up to" them, so autonomy is also incompatible with indeterminism.

Perhaps the most powerful autonomy-skeptical argument is Strawson's *impossibility argument* (Strawson 1986, 2016). He argues that autonomy is impossible—independently of considerations about whether determinism is true—because every choice is conditioned by whatever mental state we are in at and/or 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 previous mental states, back to our earliest mental state(s). But if we can never be in an unconditioned mental state, we can never be ultimately responsible for any mental state we are in, nor for any choice we make that is conditioned by that mental state. Thus, autonomy 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 autonomy and its compatibility with determinism, indeterminism, and the conditioning of mental states. I argue that meditation and related volition-regulating disciplines that lead Buddhists toward liberation increase autonomy-related abilities in the soft compatibilist sense, and that an analysis of these abilities provides the basis for a Buddhist counter to these skeptical arguments. The Buddhist meditative path is designed to eliminate what Buddhist theory sees as the illusory belief in an ontologically substantive ego/self, and with it the unregulated expression of ego-volitional preferences seen as fueling the illusory sense of ego/self. However, the unregulated expression of ego-based volitions is, from a Western perspective, thought to be partly constitutive of autonomy. Western thought prizes the autonomy-related ability to act on one's desires and preferences, do as one pleases, and make choices that are *up to* one whenever alternatives pull one in different directions. Thus, whereas Buddhism generally eschews the idea of catering to ego-volitional preferences, Western philosophy has struggled to square these autonomy-related abilities with determinism (in earlier times, to square them with God's foreknowledge, and more recently with indeterminism). The closest to this that Buddhism has struggled to explain is the problem of how volitional (karmic) formations can accrue to, or how reincarnation is possible for, individuals that are thought to lack substantive selves.

Historically, Buddhism has not been concerned with how individuals may be *properly held* morally responsible, for a number of inter-related reasons. First, because—though individual actors are considered causally (karmically) responsible, and though karma has positive/negative connotations consistent with a kind of consequentialism in which suffering should be reduced and happiness promoted—karmic consequences are not seen as retributive, but purely causal (Goodman 2009). Second, because blame of self/others are sometimes treated differently. In the *Bodhicāryāvatāra*, Śāntideva advises us to see others' actions as ultimately impersonally caused (like bile production), but to view our own actions as subject to self-regulation for soteriological purposes (Shantideva and Padmakara 2009, 6:22-32), which self-regulative efforts avowed Buddhists accept as ethically binding them, particu-

larly in later (Mahāyāna) Buddhism, to cultivate the *pāramitās*, six ethical character perfections (generosity, moral disposition, perseverance, energy, meditative attainment, and wisdom), as well as the *brahmavihāras*, the four divine abodes (loving-kindness, compassion, sympathetic joy, and equanimity). This framework does not fit neatly into Western metaethical categories, but focuses on perfecting ethical cultivation within the soteriological framework of the Buddhist path (Garfield 2015). Third, because Buddhist philosophy restricts warrants for philosophical inquiry to what is soteriologically instrumental, and the issue of blame and its problematic relationship with impersonal causal processes is not considered warranted (Gowans 2016). Fourth, because Buddhism sees agency as ultimately impersonal and its illusion as the cause of suffering, and is thus more concerned with eliminating ego-volitional agency in favor of selfless/altruistic volition than with defending it.

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 the problem of autonomy (Adam 2010; Coseru 2016; Davis 2016; Federman 2010, 2016; Friquegnon 2016; Garfield 2014; Gier and Kjellberg 2004; Gómez 1975; Goodman 2009, 2016; Griffiths 1982; Harvey 2007, 2016; Hyland 2014; Meyers 2010, 2016; Rāhula 1974; Repetti 2016; Siderits 1987, 2008; Story 1976; Tuske 2013; Wallace 2011). 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 (Blackmore 2014; Davis and Thompson 2015; Flanagan 2011; Strawson 1986, 2016; Thompson 2015). Toward this end, I defend the idea that Buddhist thought can ground a moral-responsibility-entailing autonomy against all four major skeptical arguments, but let us first sketch the *autonomist possibilism* that represents the target of these skeptical arguments.

Classical compatibilists, like Hume, argued that autonomy is compatible with determinism because it is not the fact *that* our actions are caused that matters here, but *how* they are caused. Classical compatibilists offer divergent accounts of how our actions are supposed to be appropriately caused, but typically these accounts are articulated in *negative* terms—involving the absence of freedom-undermining conditions. For

example, if not coerced, constrained, compulsive, the result of brainwashing, torture, or manipulation, all freedom-undermining conditions, then *that they are caused* need not undermine the claim that they may be sufficiently free in the moral-responsibility-entailing sense. *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.

Though much in Buddhism and in thinking about autonomy exceeds it, I focus on what fits within a naturalistic framework, to occupy the more parsimonious position. Central Buddhist ideas posing an immediate problem for the notion of autonomy are: *impermanence*, *mereological reductionism* or *eliminativism* (the partial or complete denial of partite wholes), *dependent origination*, and *insubstantiality* or *emptiness*. Each of these ideas implies *anātman* (non-self), denial of the (non-Buddhist Indian) *ātman* (the unchanging, immaterial self/soul). Impermanence characterizes all conditioned phenomena (*Dhammapada* 277-279): nothing in form (spatiotemporal) lasts unchanged 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.

A whole is at best just a way of designating a configuration of parts, but does not exist independently of the parts, nor is it identical to them (for they are many and it is one): it is merely a conceptual imputation; see *Milindapañha*, *Mūlamadhyamakakārikā*, and Siderits (2003). Thus, the 'self' is just a convenient designation for the psychophysical series. Some Buddhists are reductionists, others eliminativists, but all reject the ultimate or fundamental reality of self. Buddhist reductionism/eliminativism (depending on the interpretation) is classically depicted in the *Milindapañha*, when Buddhist monk Nagasena argues with Bactrian King Milinda (thought to be Menander) that just as the 'chariot' exists only as a conventional designation for its pragmatically configured parts, but is neither identical to nor additional to them, so too are all part/whole relations, including Nagasena and Milinda.

According to the doctrine of dependent origination (*pratītya samutpāda*), which *resembles* determinism, because all conditioned phenomena originate in dependence on other conditioned phenomena, no conditioned phenomenon exists autonomously. The Buddha stated, "When that is present, this comes to be; on the arising of that, this arises. When that is absent, this does not come to be; on the cessation of that, this ceases." (*Majjhima Nikāya* 1.262 ff.; *Saṃyutta Nikāya* 2.28) An organism, say, cannot exist independently of extra-organismic, eco-system-dependent conditions.

Since no conditioned phenomena exist independently, all conditioned phenomena and beings are metaphysically *insubstantial* or *empty*, lacking any independent ontological/existential essence, status, or *intrinsic nature*. Western philosophers might interpret this as a Buddhist claim to the effect that there are no natural kinds, the self included. 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 or of tasting lemonade. The self lacks substantive reality on all these grounds.

Further, the Buddha rejected the idea that any of the (mereological) 'aggregates' that compose us (body, volition, cognition, feeling, etc.) constitute a self, because if such an aggregate did, it could will itself—without mediation—into a non-afflicted state, but no part of us can (*Anattalakkhaṇa Sutta*; Adam 2010). Just as Western philosophers have objected that the Cartesian mind—being immaterial—cannot interact with any physical brain/body, exposing the incoherence of the notion of immaterial mind, Buddhist philosophers, principal among them Nāgārjuna, objected that the unchanging nonphysical *ātman* would be incapable of doing anything (*Mūlamadhyamakakārikā*; Federman 2010). These arguments are mutually confirming: whatever might constitute self cannot be causally inert, therefore cannot exist independently, and thus must be dependently originated, momentary, partite, and empty, and vice versa.

Given the Buddhist *anātman* doctrine it seems Buddhism would equally deny the reality of autonomy. For Goodman (2009, 2016), a non-agent/non-self cannot

be autonomous. 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 (Repetti 2012b). Those who deny partite wholes like chariots—the classic Buddhist analogy—or selves may be pressed to deny the parts of parts, and so on, all the way down (Siderits 2003, 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 (Repetti 2012a). If this is an argument against the reality of self, it is also an argument against the reality of chemical bonds, hands, chairs, Buddhas, and just about everything.

Thus, attempts to rest negative conclusions about autonomy on the rejection of wholes substitute one mystery for another, and beg the question, because their premise is more problematic than their 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 fact that plants cannot exist independently of the photons that fuel their necessary photosynthesis does not entail that *there are no plants*, but only that there are no ontologically autonomous plants—plants that exist independently of plant-exogenous conditions. 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 constantly replacing. 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*. Simplifying *greatly*, Buddhism acknowledges the conventional existence of plants, money, sentient beings, and so on, but denies that such entities exist in ultimate reality. (This distinction is interpreted differently within major doctrinal divisions of Buddhism, too complex

to go into here, but analogous to the way empiricism or phenomenology are understood differently within different Western traditions. For an excellent introduction to the major doctrinal views on the distinction based on authoritative Buddhist texts, see Thakchoe (2011); for an in-depth analysis, see Thakchoe 2007.) Arguably, the *two-truths* distinction maps loosely onto the difference between mereological reductionism and mereological eliminativism: conventional reality, on this view, involves a form of mereological reductionism about wholes, in that wholes sort of exist, somewhat, pragmatically, but only because they supervene on or reduce ultimately to things that really do exist at the micro-level, and ultimate reality, on this view, involves a form of mereological eliminativism about wholes, in that they don't really exist at all (Siderits 2008). Joining these, we may say conventionally bridges exist, but ultimately nothing exists independently of our conceptual designation of the dependently-originated, momentary, insubstantial/ empty conditioned phenomena the aggregation of which we pragmatically identify as 'bridges.'

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 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 2003).

Despite its promising appearances, the two-truths doctrine (hereafter, "2Ts") admits of divergent, complex interpretations (Thakchoe 2007, 2011), and seems more problematic than autonomy does. Thus, appeal to 2Ts is question begging, for if 2Ts is to ground a negative view of autonomy, as a premise 2Ts requires greater support than the conclusion it purports to support (Repetti 2012a, 2012b). For 2Ts not only denies the ultimate reality of autonomy, but also of hands, money, and just about everything else—except, on one reading (accepted only by pre-Mahāyāna Buddhists), the ultimate reality of *dharmas* (Sanskrit; Pāli: *dhammas*, micro-phenomena): impartite atomistic psychophysical tropes like whiteness, sweetness, volition, and the like (Siderits 2003, 2008), which en-

tities are, from a Western philosophical and scientific perspective, of a dubious nature (Repetti 2012a). At the same time, 2Ts is analogous to the scientific bifurcation of such pairs of folk/scientific concepts as *water/H²0*, *weight/mass*, and *heat/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 self-regulation seems hardly eliminative. I see no bar in Buddhism to the reductionist interpretation, which I have developed elsewhere (Repetti 2012a, 2012b) and sketch below.

Challenges to Autonomy from Causation

The four main skeptical arguments discussed above may be grouped together with a variety of related skeptical challenges to autonomy 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 produced, conditioned, or influenced in some way or another it cannot be free. The consequence argument may be the most straightforward challenge from causation because it asserts that if putatively free choices are caused by prior states of the universe in accordance with deterministic laws, they cannot be free, because alternatives are never available to agents to choose otherwise than whatever choice the earlier world-state determined. The manipulation argument counts as a challenge from causation because it asserts that there is no principled difference between manipulation of an agent's beliefs, desires, or will by anything extrinsic to the agent, which renders the agent's choices unfree, on the one hand, and determinism, on the other hand. The randomness argument counts as a challenge from causation because it asserts that even if the will is indeterministically caused it is not free, because if choices are indeterministic, their random occurrence is beyond the agent's control. The impossibility argument counts as a challenge from causation because it asserts that the agent's mental states condition her choices, so they cannot be free, and conditioning involves some kind of causal influence, however unspecified. Illusionist arguments count as challenges from causation because they rest on empirical findings that suggest causal deviance beneath various cognitive abilities associated with autonomy (e.g., unconscious biases, post-facto confabulation, etc.), implying that our conscious states are causally impotent in the production of our choices.

Optimist's and Pessimist's Dilemmas

Autonomy pessimists pose a certain dilemma for autonomy 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 autonomy. 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; thus it is an accident which future emerges from prior world-states, akin to the outcome of a coin toss.

Consider a dilemma faced by the optimist, from a conjunction of the consequence argument's and/or the manipulation argument's deterministic challenge and the argument from randomness, the *optimist's dilemma*:

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

This argument is the basis for hard incompatibilism, all components of which have been discussed above. However, as I have argued (Repetti 2010a), an opposite dilemma may be posed for the pessimist by retaining claim 1, negating and transposing the consequents across 2 and 3, yielding the opposite of claim 4, the *pessimist's dilemma*:

1. Either determinism or indeterminism is true.
2. If determinism is true, choices are not random, thus agents *can* claim to author them, and may possess autonomy.
3. If indeterminism is true, agents *can* bring about

things not already determined (and satisfy criteria that make a difference), and may possess autonomy.

4. Thus, either way, autonomy is possible.

This argument is the basis for soft compatibilism. In support of its premises 2 and 3, I need only refer to 2 and 3 in the optimist's dilemma, transposing their conditionality, and to the works of other philosophers. Regarding premise 3, Kane (2002), Mele (1995), and others have given good reasons to think indeterminism may be situated in parts of the deliberative process that would *enhance* autonomy; e.g., Kane constructs a model in 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 premise 2, Frankfurt (1971), Dennett (1984), Flanagan (2002), Fischer (2006), 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.

Dispositions, Counterfactuals, Possibilities, and Actualism

Let us now challenge hard determinism. Just as salt is soluble (in liquid), an agent is able to raise his hand (voluntarily). Consider a 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.

Hard determinists may object that salt cannot bring it about that it is placed in water. 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 subjunctives of the form, "if this salt *were* submerged in liquid, then it *would* dissolve." Thus, it is *possible* salt dissolves, *if* certain conditions arise, *even if* salt is never exposed to them, and never *actually* dis-

solves; alternately, there is a possible world in which salt is thus exposed and 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 *just is* an abstraction from all counterfactual-supporting generalizations or laws; for example, if A's necessarily cause B's, this supports the counterfactual *if there had been an A, then there would have been a B*, or the subjunctive *if there were an A, then there would be a B*. (Here I note, but set aside, worries about nearby possible worlds in which blockers or masks render it false that B *accompanies* A, even if A *causes* B, since such complications don't bear on the present point.) *Deterministic actualism* is thus oxymoronic; by default, *deterministic counterfactualism* is tautologous. Thus, the following approximation/sketch of a 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 sketch *the counterfactual principle of autonomy* ("CPA"). CPA-type principles may be used to identify a determinism-friendly model of responsible agency (Repetti 2010a).

Unlike salt, which is helpless regarding whether it meets its solubility condition (liquid immersion), agents can bring about conditions needed to increase what Buddhism identifies as *mental freedom* and thus can cultivate abilities that increasingly satisfy CPA conditions. Buddhism specifies how we may do so in its many prescriptions for approaching *nirvāṇa* (Sanskrit; Pāli: *nibbāna*, total mental freedom), by cultivating, through meditative disciplines, true beliefs, wholesome volitions, mindfulness, mental focus, mental stability, detachment, phenomenological insight into the triggers of our volitional processes, mental equilibrium, etc., and by decreasing such things as false beliefs, unwholesome (ego-based) volitions,

mindlessness, mental disturbances (anger, jealousy, greed, hatred, craving), attachment, etc. Insofar as an agent follows the prescription he will attain mental freedom, which admits of degrees, and—this is the main point of this paper—insofar as a being increasingly approximates *nirvāṇa* she increasingly satisfies CPA in her choices, thereby exhibiting increasingly effective autonomy.

Meditative practice is at the core of the development of these skills. Buddhism classifies meditation types—simplifying greatly—as involving some concatenation of *one-pointedness* and *mindfulness*. These terms have technical meanings in Buddhism that are sometimes conflated in English; in fact, many traditional Buddhists (Purser, Forbes, and Burke 2016; Purser and Loy 2013; Sharf 2012) object to the widely popular Western extraction of the ethically-neutered 'mindfulness' as 'bare attention' successfully adapted to psychotherapeutic uses in the mindfulness-based stress reduction and mindfulness-based cognitive therapy movements spearheaded by Kabat-Zinn (1991), which have mushroomed interest in 'mindfulness' from the military to leadership, education. (For opposing views, see Brown et al. (2015), Federman (2016), Hyland (2014), Repetti (2016c).) Even philosophy is on board (Davis 2016; Davis and Thompson 2015; Garfield 2015; Repetti 2010b, 2016a, 2016b; Strawson 1986; 2016; Thompson 2015).

Mindfulness is the translation for *sati* (Pāli; Sanskrit: *smṛti*), which also suggests *remembering* and *familiarity*, but as a technique refers to the ability to attend non-conceptually to the experiential/phenomenological features of the object of focus, thought to lead to the deepest (Buddhist) metaphysical insights, i.e., impermanence, interdependence, and insubstantiality/emptiness. *Insight* is the translation for *vipassanā* (Pāli; Sanskrit: *vipaśyanā*), although popular Anglophone usage loosely translates it as *mindfulness*. *One-pointedness* is the translation for *samadhi* (Pāli, Sanskrit), the ability to maintain focus, thought to lead to tranquility, 'mental quiescence' 'calm abiding,' (Sanskrit: *śamatha*). Meditative *discipline* is also the translation for *samādhi*, which includes both mindfulness and one-pointedness as well as effort (Lutz, Dunne, and Davidson 2007)

These practices also cultivate patience, detachment, greater distance or 'elbow room' between stimuli/impulses and responses, and thus increased self-control (Repetti 2010a; 2010b; Harvey 2007; Federman 2010). These skills are cultivated within a philosophical/soteriological system designed to reduce mental bondage and increase mental freedom. Those three insights noted above arose upon the Buddha's meditative penetration into the phenomenology of experience: impermanence (all is momentary), interdependence (everything is dependent on everything that contributed to its existence), and insubstantiality/emptiness (everything lacks metaphysical essence/substance). The *nirvāṇa*-generating insight that follows is that there is no metaphysically substantive self. As one assimilates the unreality of self, ego-volitions diminish.

Buddhist versions of CPA might express such ideas as that if agent A were thoroughly mindful of A's volitions, A would be less driven by them 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), A would have displaced V1's force. 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 self-regulation (agency/autonomy).

Even if responsible agency is not afforded *intrinsic* value in Buddhism (because Buddhism seeks to move away from reactive attitudes and the idea of an autonomous controller of actions), these considerations show that the responsible agency adumbrated here has *instrumental* (soteriological) value. Thus, though Buddhism avoids abstract philosophical discussions that lack soteriological warrants, this discussion is warranted.

Strawson's Impossibility Argument

As noted earlier, *influence* counts as a challenge from causation. Though Strawson is a kind of compatibilist (Repetti 2011), his impossibility argument rests on *influence*, posing a more general challenge than 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 it a *function* of that state and thus not ultimately free in the responsibility-entailing sense.

Strawson argues that the only way one could be un-

influenced and thus ultimately responsible for 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* (from no conditions), a kind of *causa sui* or self-created being, which is logically impossible (to create oneself one must exist prior to one's creation).

Strawson's argument may be taken, *arguendo*, to gloss the other causal challenges. For, whether one's mental state is deterministically, indeterministically, manipulatively, or unconsciously/neurally generated, it remains that *something influences* one's mental state in the choice-moment. Thus, the impossibility argument *subsumes* most if not all skeptical, pessimistic arguments discussed above, since *how* one's choice-moment-mental-state is influenced is irrelevant to the fact that *it is* influenced.

The Buddhist rebuttal to this version of the impossibility argument constitutes an argument—not a proof—for possibilism, and a potential rebuttal to many of the other skeptical, pessimist arguments and positions, excepting any that might have relevantly unique features. (Clarke (2005) has a rebuttal to Strawson's impossibility argument, for example, that likely does not apply to the manipulation argument.) Again, though Buddhism has historically ignored the autonomy issue and some contemporary Buddhist scholars prefer to, that Buddhism has the resources to formulate a possibility argument that can rebut many, most, if not all of these challenges from causation is reason alone not only to consider it, but to take it seriously.

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

In my view, a Buddhist conception of autonomy may be formulated negatively, as a kind of 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. In Buddhism, all ego-based volitions exemplify this relationship between ignorance about 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 influences may be viewed usefully through the lens of Frankfurt's distinction between first-order desires (for experiences or things) and higher-order desires or metavolitions (that have other volitions or desires as their objects, or pro- and con-attitudes towards certain first-order desires).

The *will* of a person is typically volitionally/metavolitionally structured or hierarchical, for Frankfurt, whereas the will of an infant or nonhuman animal is typically unstructured, entirely first-order. *Freedom of will*, for Frankfurt, is having the sort of will one wants, such that the first-order desires one meta-volitionally approves are the only ones that succeed in action, whereas *weakness of will* occurs when metavolitions fail to regulate volitions. Chronic cases involve 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 (Repetti 2010a), particularly insofar as they fit nicely with a certain Buddhist conception of freedom. Note that I am not arguing that Buddhism endorses *Frankfurt's theory* of autonomy, but only that it may adapt *Frankfurt's distinctions* to fruitful ends.

Buddhist meditators are increasingly able to form the sort of effective, hierarchically structured, mental-freedom-oriented wills they want, through long, disciplined effort. Buddhist meditative training involves long-term discipline with *one-pointedness* and *mindfulness*, an introspective analogue of intense athletic (proprioceptive) training of the body or of scientific (exteroceptive) training of analytic observation that 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 (Thompson 2015). 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 (Harvey 2007; Federman 2010). 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 skilled at fostering those cognitive/conative influences that are freedom-enhancing, and diminishing those that are bondage-enhancing (Wallace 2011). 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 *theory* notwithstanding. Nonetheless, let's review the problem with Frankfurt's theory, and show how easily Buddhism can circumvent it.

Critics object that the theory implies it is a higher-level volition, say, L2, that renders the lower-level volition, L1, free (when L2 endorses L1), but to render L2 free it must be endorsed by L3, and so on, which self-iterative formula generates a regress (Repetti 2010a; Tuske 2013). Even if this is fatal to Frankfurt's *theory* (which I doubt), it is not fatal to my use of his *distinctions* within the CPA framework: the Buddhist model needs no iterative formula, but only the claim that metavolitional CPAs correlate increases in mental freedom with increases in responsible agency, avoiding any regress (Repetti 2010b).

The Fallacy of the Heap

The Frankfurt-informed Buddhist model of responsible agency formulated above may be used to undermine Strawson's impossibility argument, although possibly not directly targeting what Strawson targets, as I'll explain. 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* and other texts (Bodhi 1995; Buddhaghosa and Nanamoli 2003), for example, which texts specify the stages on the path to *nirvāṇa*, with extremely fine-grained phenomenologically-grounded analyses of various otherwise experientially coarse-grained mental states that involve differentiations between psychophysical tropes counted in the trillions per blink of an eye (*Vissudhimagga*).

Skilled in introspective phenomenological investigation into the dynamics of volition, *āryas* (adept 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 *āryas* toward the better 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 specify the number of grains of sand required for heaps, but surely there are small 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, Strawson's argument implicitly differentiates between conditioning influences of two quantities, *0* and *0+n* (where *n* is a non-zero or positive quantity), and parses freedom and unfreedom, respectively. Buddhism offers the possibility of transitioning from and thus escaping unfreedom (*0+n* influence) and attaining total mental freedom (*0* influence) upon the attainment of *nirvāṇa*, which may be defined for our purposes as liberation from all ego-volitional impulse and cognitive delusion—*all mental-freedom-undermining 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*, so one way to diagnose Strawson's error would be to say he wrongly presupposes that the only way to attain *0* influence is to be a *causa sui*. Alternatives include *nirvāṇa*, being created with libertarian free will by God, being created somehow analogously by an expert simulated-world programmer (such that all your programming leaves it open how your first choice emerges), etc. *Nirvāṇa* is a possibility model for what the impossibility argument says is impossible. It also reveals that argument commits a sorites fallacy.

I noted a qualification in the beginning of this section about the targets of Strawson's impossibility argument and of the Buddhist response, suggesting possible cross-purposes. Strawson specifically targets non-naturalistic, *ultimate*-moral-responsibility-entailing conceptions of agency, but does not target naturalistic, conventional conceptions of moral-responsibility-entailing agency that may be described as proximal, relative, or compatibilist. Since the Buddhist alternative on proposal here consists of some version of the latter conception, arguably Strawson's impossibility argument does not threaten the Buddhist alternative, in which case one might object that the Buddhist argument here misses the mark. True, insofar as Buddhism offers a naturalistic alternative, it is abstractly immune from arguments that target non-naturalistic conceptions (e.g., the *causa sui*). That—and how—Buddhism circumscribes such arguments is no doubt philosophically interesting.

However, Strawson's impossibility argument has collateral damage for Buddhism insofar as it implies that *nirvāṇa*—freedom from all present/prior conditioning—is impossible. Thus, Buddhism is warranted in responding to this collateral damage. Indeed, I take it that it is 'collateral *non*-damage' and a virtue of the Buddhist defense of *nirvāṇa* that it undermines one of the most powerful anti-autonomy arguments.

Dharma-responsive Responsible Agency

The Buddhist teachings, path, or way may be referred to as the *Dharma* (Sanskrit; Pāli: *Dhamma*), which term more generally denotes the soteriologically relevant dimension of the truth (according to Buddhist understanding). If an agent is sufficiently dharmic—sensitive and responsive to the truths of impermanence, insubstantiality, and impersonality as well as to the suffering of sentient beings—she is what I have described as *Dharma-responsive* (Repetti 2010b). *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 the extent to which agents exhibit determinism-friendly responsible agency.

Semi-compatibilists separate moral responsibility from autonomy: they are incompatible (*hard*) determinists about autonomy because they think it requires indeterminism, but compatible (*soft*) determinists about moral responsibility because they think reason-responsiveness suffices for it and is compatible with determinism. Reason-responsiveness may be compared 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 (e.g., a rusted one) is neither. Likewise, a rational being recognizes relevant reasons in varying contexts and responds to them appropriately in choice and action.

Wolf, Fischer and other compatibilists and semi-compatibilists reason differently to the shared conclusion that even if agents lack the kind of autonomous ability required to do otherwise under identical conditions (impossible in a deterministic world, but possible in an indeterministic world or under a dualistic model), insofar as they are moderately reason-responsive, they have enough of a compatibilist kind of autonomy that renders them morally responsible for their actions, precisely because they are able to recognize and respond appropriately to moral reasons—they are functional moral agents, able to regulate volitions for moral reasons (equivalent to metavolitions about volitions).

Dharmic reasons are a species of reasons; *Dharma*-responsiveness is a species of reason-responsiveness.

Thus, just as reason-responsiveness grounds compatibilist moral agency, so does *Dharma*-responsiveness. This is a logical point, but substantive features of *Dharma*-responsiveness establish the same conclusion on extra-logical, Buddhism-intrinsic grounds. For a *Dharma*-responsive agent is independently sensitive to a network of moral reasons that concern whether entities will suffer and that exclude biological and related attributes (kin, gender, species, etc.), rendering them altruistic reasons.

Thus, *Dharma*-responsive agents are not only responsible agents, but at a level most consider supererogatory. Responsible agency here may be construed as causal/metaphysical responsibility, not merely utilitarian in the sense critics intend by the idea that fostering attitudes, norms, and institutions of *holding* people responsible (possibly by Skinnerian means)—even though, metaphysically, they are not responsible—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 possesses what may be considered Buddhist autonomy.

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 autonomy. Nonetheless, despite its intrinsic unimportance, this form of autonomy 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* (a concept explored in depth in Repetti 2016; 2016b).

Autonomous Non-agents? The Deflationary Self

This model of responsible agency is consistent with deflationary models of the self, as articulated by Siderits (2003), Strawson (1986), Flanagan (2002), Thompson (2015), 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. This ultimately impersonal person-series/set of (skillfully self-regulating) processes Buddhists refer to conventionally as an *ārya*, who through practice cultivates effective metavolitions, is highly (causally) responsible for shaping and sculpting the sort of self-process the *ārya* becomes over time, and the more she does, the greater her (morally) 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—conceived by itself as—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

Socratic wisdom consists in awareness of one's ignorance, of the difference between what one does and doesn't know. Buddhism, philosophy, and science share this aim of uncovering ways in which we are subject to ignorance/illusion, in favor of a more ultimate reality (Blackmore 2014). 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. It seeks to bring practitioners into greater awareness of these delusions. In these regards, it aspires to enacting and embodying Socratic wisdom.

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 also fares well against other powerful forms of autonomy pessimism, such as the consequence and manipulation arguments. Let's review the consequence argument to segue into the more challenging manipulation argument, which builds on the consequence argument idea that determined behavior cannot be free.

The consequence argument entails 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, lacks autonomy, and cannot be morally responsible (van Inwagen 1975). But neo-compatibilists like Frankfurt (1971), Dennett (1984), and Flanagan (2002) argue plausibly, along divergent lines, that causation is consistent with some version of responsible agency—lines similar to those about metavolitional/volitional regulation, reason-responsiveness, and so forth. This is likely why incompatibilists tend to rely more heavily, as many, like Pereboom (2001), 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 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 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 C, circumstances in the agent's environment are manipulated so as to lead the agent to make that same choice; and, in 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 (responsible-agency-undermining) 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, in the other via deterministic means), all versions 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 (Vihvelin 2011). This is considered by many to be the strongest argument against compatibilism. What can the Buddhist say in response?

Enlightened Brains in Vats

Suppose the *ārya* has satisfied sufficiently many Buddhism-specific CPA-type principles, whatever they may be, rendering her significantly metavolitionally self-regulating, *Dharma*-responsive, etc., thus exhibiting a robust Buddhist form of autonomy. Suppose further that *ārya* was manipulated by agent-extrinsic means: conjure up whichever type of manipulative scenario suits your 'phi-fi' imagination (neural implants, cult brainwashing, futuristic bionic/cyborg brains designed in Skinnerian societies—whatever). (I coined the neologism *phi-fi* for *philosophical fiction*, akin to *sci-fi* (Repetti 2010a).) 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 than the more generic compatibilist can. Suppose the neural correlates of a robust form of autonomy or even of *nirvāṇa* could be generated by neuroscientific manipulation. Would that undermine them? *Absolutely not!* With manipulative enemies like that, one would not need friends. 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, 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. If only it was so easy!

The generic compatibilist may need to differentiate between manipulation and determinism, insofar as such a theory is committed to backward-looking or historical considerations (i.e., to identify features of its causal history that explain how it came to act freely). For the autonomous agent on that 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—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 (Federman 2010).

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 their origins. The *ārya*, having reached an advanced stage in practice (by whatever 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 manipulation. In fact, certain Buddhists advocate the use of fear and other manipulative devices as *skillful means* (a Buddhist doctrine approving otherwise immoral means to soteriological ends) 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 a futuristic Buddhist 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 *ārya* 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 *ārya* 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 an *ārya* mind/brain in a vat can exhibit responsible agency and reach *nirvāṇa*. For such a minded brain 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, motor nerves, limbs, and 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, 2015).

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: nobody is left to count as the target of manipulation.

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 2014).

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 (Thompson 2015), but initial evidence along such lines is already promising (Ie, Ngnoumen, and Langer, 2014). 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 on *ārya* brains are actually helping neuroscientists identify neural correlates of conscious mental states (Thompson 2015). 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 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). Studies not connected at all with meditation but rather simply with contrasting cases of mindfulness and mindlessness confirm similar results (Ie, Ngnoumen, and Langer, 2014). 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 (somewhat paradoxically agentless) autonomy 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, responsible agency, whether their mental state at

the choice-moment is deterministic, manipulated, or conditioned.

Acknowledgements

References

• 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.

• 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. http://www.accesstoinsight.org/lib/authors/bodhi/abhiman.html (accessed July 24, 2014).

• Brown, K.W., J.D. Creswell, and R.M. Ryan. 2015. *Handbook of mindfulness*. New York: Guilford Press.

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

• Caruso, Gregg, ed. 2013. *Exploring the illusion of free will and moral responsibility*. New York: Lexington Books.

• Clarke, Randolph. 2005. On an argument for the impossibility of moral responsibility, *Midwest Studies in Philosophy* 29, 13-24. http://dx.doi.org/10.1111/j.1475-4975.2005.00103.x

• Coseru, Christian. 2016. Freedom from responsibility: Agent-neutral consequentialism and the bodhisattva ideal. In *Buddhism and free will: Agentless agency*, edited by Rick Repetti, ch. 5. London: Routledge.

• Davis, Jake. 2016. Reactive attitudes without reactivity: Does mindfulness take us beyond moralizing? In *Buddhism and free will: Agentless agency*, edited by Rick Repetti, ch. 17. London: Routledge.

• Davis, Jake H. and Evan Thompson. 2014. Developing attention and decreasing affective bias: toward a cross-cultural cognitive science of mindfulness. In *Handbook of mindfulness*, edited by K.W. Brown, J.D. Creswell, and R.M. Ryan. New York: Guilford Press, 42-62.

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

• Federman, Asaf. 2016. Free will in contemporary mindfulness-based programs. In *Buddhism and free will: Agentless agency*, edited by Rick Repetti, ch. 19. London: Routledge.

• Federman, Asaf. 2010. What kind of free will did the Buddha teach? *Philosophy East and West* 60, 1-19. http://dx.doi.org/10.1353/pew.0.0086

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

• Flanagan, Owen. 2011. *The bodhisattva's brain: Buddhism naturalized*. Cambridge: MIT Press.

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

• Frankfurt, Harry. 1971. Freedom of the will and the concept of the person, *Journal of Philosophy* 68, 5-20. http://dx.doi.org/10.2307/2024717

• Friquegnon, Marie. 2016. Buddhism and free will. In *Buddhism and free will: Agentless agency*, edited by Rick Repetti, ch. 13. London: Routledge.

• Garfield, Jay. 2015. *Engaging Buddhism: Why it matters to philosophy*. New York: Oxford University Press.

• Garfield, Jay. 2014. Just another word for nothing left to lose: freedom of the will in Madhyamaka. In *Free will, agency and selfhood in Indian philosophy*, edited by Matthew Dasti and Edwin F. Bryant, 164-85. New York: Oxford University Press. Reprinted in Repetti 2016. http://dx.doi.org/10.1093/acprof:oso/9780190204334.001.0001

• 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. http://www.smith.edu/philosophy/docs/garfield_brain.pdf (accessed August 24, 2015).

- 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.

- Ginet, Carl. 1966. Might we have no choice? In *Freedom and determinism*, edited by Keith Lehrer, 87-104. New York: Random House.

- Gómez, Luis O. 1975. Some aspects of the free will question in the Nikāyas, *Philosophy East and West* 25, 81-90. http://dx.doi.org/10.2307/1398436

- Goodman, Charles. 2016. Uses of the illusion of agency: Why some Buddhists should believe in free will. In *Buddhism and free will: Agentless agency*, edited by Rick Repetti, ch. 2. London: Routledge.

- Goodman, Charles. 2009. *Consequences of compassion: an interpretation & defense of Buddhist ethics*. New York: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195375190.001.0001

- Gowans, Christopher. 2016. Why there was no 'problem of freedom and determinism' in Buddhism. In *Buddhism and free will: Agentless agency*, edited by Rick Repetti, ch. 1. London: Routledge.

- Griffiths, Paul J. 1982. Notes toward a Buddhist critique of karmic theory, *Religious Studies* 18, 277-291. http://dx.doi.org/10.1017/S0034412500014128

- Harvey, Peter. 2016. Psychological versus metaphysical agents: a Theravāda Buddhist view of free will and moral responsibility. In *Buddhism and free will: Agentless agency*, edited by Rick Repetti, ch. 10. London: Routledge.

- 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. http://dx.doi.org/10.1558/bsrv.v31i1.125

- Ie, Amanda, Christelle T. Ngnoumen, and Ellen Langer, eds. 2014. *The Wiley Blackwell handbook of mindfulness*. Wiley-Blackwell.

- Kabat-Zinn, Jon. 1991. *Full catastrophe living: using the wisdom of your body and mind to face stress, pain, and illness*. Delta Trade Paperbacks.

- 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. http://dx.doi.org/10.1017/CBO9780511816789.020

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

- Meyers, Karin. 2016. Relating Vasubandhu's philosophies of mind and action to the contemporary problem of free will. In *Buddhism and free will: Agentless agency*, edited by Rick Repetti, ch. 12. London: Routledge.

- 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. http://dx.doi.org/10.1017/CBO9780511498824

- Purser, Ron, David Forbes, and Adam Burke, eds. 2016. *Handbook of mindfulness: Culture, context and social engagement*. Springer Publishing.

- Purser, Ron and David Loy. 2013. Beyond McMindfulness, *Huffington Post*. July 1, 2013. http://www.huffingtonpost.com/ron-purser/beyond-mcmindfulness_b_3519289.html (accessed August 24, 2015).

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

- Repetti, Rick. 2016a. Why there ought to be a Buddhist theory of free will. In *Buddhism and free will: Agentless agency*, edited by Rick Repetti, ch. 5. London: Routledge.

- Repetti, Rick. 2016b. Buddhist meditation, mental freedom, and free will. In *Buddhism and free will: Agentless agency*, edited by Rick Repetti, ch. 20. London: Routledge.

- Repetti, Rick. 2016c. Meditation matters: Replies to the anti-McMindfulness bandwagon. In *Handbook of mindfulness: Culture, context and social engagement*. Springer Publishing, edited by Ron Purser, David Forbes, and Adam Burke.

- Repetti, Rick, ed. 2016. *Buddhism and free will: Agentless agency*. London: Routledge.

- Repetti, Riccardo. 2014. Recent Buddhist theories

of free will: Compatibilism, incompatibilism, and beyond, *Journal of Buddhist Ethics* 21, 279-352.

• Repetti, Riccardo. 2012a. Buddhist reductionism and free will: Paleo-compatibilism, *Journal of Buddhist Ethics* 19, 33-95.

• Repetti, Riccardo. 2012b. Buddhist hard determinism: No self, no free will, no responsibility, *Journal of Buddhist Ethics* 19, 130-97.

• Repetti, Riccardo. 2011. Free-will-possessing and enlightened Buddhist brains in vats. Panel Presentation. Contemporary Perspectives on Buddhist Ethics. Columbia University. October 7, 2011. http://www.cbs.columbia.edu/buddhist_ethics/panel-five.html (accessed August 24, 2015).

• Repetti, Riccardo. 2010c. Earlier Buddhist theories of free will: Compatibilism, *Journal of Buddhist Ethics* 17, 279-310.

• Repetti, Riccardo. 2010b. Meditation and mental freedom: a Buddhist theory of free will, *Journal of Buddhist Ethics* 17, 166-212.

• Repetti, Riccardo. 2010a. *The counterfactual theory of free will: A genuinely deterministic form of soft determinism.* Saarbrücken, Germany: LAP Lambert Academic.

• 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.

• Sharf, Robert H. 2012. Varieties of mindfulness. mindfulness or mindlessness: traditional and modern Buddhist critiques of 'bare awareness.' Lecture, Advanced Study Institute, McGill University. June 3-5, 2012. https://www.youtube.com/watch?v=c6Avs5iwACs (accessed August 24, 2015).

• Siderits, Mark. 2008. Paleo-compatibilism and Buddhist reductionism, *Sophia* 47:1, 29-42. Reprinted in Repetti 2016.

• Siderits, Mark. 2003. *Personal identity and Buddhist philosophy: Empty persons.* Aldershot, UK: Ashgate.

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

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

• Strawson, Galen. 2016. Free will and the sense of self. In *Buddhism and free will: Agentless agency*, edited by Rick Repetti, ch. 6. London: Routledge.

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

• Thakchoe, Sonam. 2011. The theory of two truths in India, *Stanford encyclopedia of philosophy.* http://plato.stanford.edu/archives/sum2011/entries/twotruths-india/ (accessed August 24, 2015).

• Thakchoe, Sonam. 2007. *The two truths debate: Tsongkhapa and Gorampa on the middle way.* Boston: Wisdom Publications.

• Thompson, Evan. 2014. *Waking, dreaming, being: self and consciousness in neuroscience, meditation, and philosophy.* New York: Columbia University Press.

• 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. http://dx.doi.org/10.1002/9781118324004.ch27

• Van Inwagen, Peter. 1975. The incompatibility of free will and determinism, *Philosophical Studies* 27, 185-99. http://dx.doi.org/10.1007/BF01624156

• Vihvelin, Kadri. 2011. Arguments for incompatibilism, *Stanford encyclopedia of philosophy.* http://plato.stanford.edu/archives/spr2011/entries/incompatibilism-arguments (accessed August 24, 2015).

• Wallace, B. Alan. 2011. A Buddhist view of free will: beyond determinism and indeterminism, *Journal of Consciousness Studies* 18(3-4): 217-33. Reprinted in Repetti 2016.

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