

Chapter 6

Breaking Good

MORAL AGENCY, NEUROETHICS, AND
THE SPONTANEITY OF COMPASSION

Christian Coseru

6.1. Introduction

Although the credit for announcing the advent of “neural Buddhism” goes to the *New York Times* op-ed columnist David Brooks (2008)—who sees it as the natural outcome of a new wave of research into the neuroscience of religious experience—it is Nietzsche who most eloquently (and disquietingly) proclaimed its arrival more than a century ago. In *On the Genealogy of Morality*, Nietzsche (2006, p. 7) speaks of the morality of compassion and of the many (by his own estimation sinister) ways it has cast around “even wider to catch even philosophers and make them ill”—the philosophers’ sickness being nothing but the symptom of a culture about to give birth to a new “European Buddhism.”

Contemporary moral psychology lacks a systematic account of compassion.¹ Nonetheless a little over a century since Nietzsche’s proleptic pronouncement, this genealogical quest for the roots of morality is giving birth to a new ethics: call it Buddhist neuroethics. As a specific domain of inquiry Buddhist neuroethics describes a constellation of moral and epistemological concerns about the exercise of practical reason in the age of brain science. As a taxonomical category, it simply functions in much the same way as Francisco Varela’s “neurophenomenology” and Patricia Churchland’s “neurophilosophy,” terms coined to designate new domains of inquiry born from the relevance and applicability of neuroscientific research to traditional issues in phenomenology and philosophy of mind. In the broadest sense of the term,

Buddhist neuroethics stands for the collective (and concerted) effort to make different aspects of moral cultivation and contemplative practice receptive to the findings and conceptual resources of neuroscience. As such, it shares many features with programs in both neurophenomenology and neurophilosophy, as well as with newer programs in neuroethics (specifically, those concerned with how understanding the human mind, and our ability to predict, influence, and even control aspects of it, affects our moral views).

This paper addresses two specific and related questions the Buddhist neuroethics program raises for our traditional understanding of Buddhist ethics: Does affective neuroscience supply enough evidence that contemplative practices such as compassion meditation can enhance normal cognitive functioning? Can such an account advance the philosophical debate concerning freedom and determinism in a profitable direction? A satisfactory answer to the first question is simply a matter of identifying the relevant empirical evidence necessary to support a Buddhist neuroethics project.² The second question does not invite a straightforward answer. The long-running debate over the compatibility of freedom and determinism has moved mainly in two directions. Those approaching the problem from a metaphysical standpoint generally argue for some version of incompatibilism, on the grounds that, if determinism is true, it is incompatible with free will, and if it is false, we are left with an indeterminism that makes free will irrelevant. Those taking an empirical approach (and thus more sensitive to the findings of cognitive science) argue for some version of compatibilism. Specifically, certain versions of neocompatibilism are regarded as capable of accommodating freedom and determinism because they see its exercise as constrained by the very causal and conditioning factors that make freedom possible. Freedom is not free (so to speak) if it is not constrained by the reasons we give for choosing one way or another. Unconstrained freedom, at least on the neocompatibilist account, is a deeply incoherent notion.

In response to the first question, I will argue that dispositions such as empathy and altruism *can* in effect be understood in terms of the mechanisms that regulate affective cognition. Not only does such understanding make a good case for causal explanation, but it also reflects the generally naturalist outlook of Buddhist moral psychology; indeed, given that dispositions and reasons have an event structure (that is, they are constituted as mental states with specific intentional content) they are also causes, or at least are causally relevant for action.³ But causal explanation is no substitute for understanding what it is about our capacity to choose that makes us moral agents. It seems as though at the most basic level choice is deeply embedded in mechanisms that regulate our capacity to discriminate and form judgments. If that is the

case, then the roots of morality lie much deeper in the structure of conscious behavior than one might think.⁴

In response to the second question, I want to claim that moral agency is a type of achievement that comes with learning the norms of ethical conduct, which are not tractable by specifically neurobiological mechanisms and processes (though, once learned, such norms would have their neural correlates when enacted). We do not hold infants morally accountable for their actions, and we recognize that childhood is at best a setting stage for the development of a moral sense. And although we admit that only adults can be considered responsible for their actions, we recognize that their comportment too reflects norms and values that are both acquired and constitutive of their moral agency. In brief, if morality is an emergent property of a certain type of socialization and not simply an adaptive trait, at least some of its features should be easily accommodated by the dynamic of social and interpersonal relations.

Nonetheless, even as a late achievement responsibility-entailing moral agency still demands that cognitive mechanisms, specifically those that regulate an individual's capacity for self-monitoring, self-control, and self-correction, are in good working order. Indeed conditions we typically associate with various psychopathies (and sociopathies) pose a challenge to this developmental account of moral responsibility. Likewise conditions associated with various forms of mental and moral cultivation suggest that introspective awareness and volitional control play a key role in modulating neuroplasticity (see Lutz, Dunne, & Davidson, 2007; Brefczynski-Lewis, Lutz, Schaefer, Levinson, & Davidson, 2008). At least for now the jury is still out on whether genetic, environmental, social, and interpersonal factors provide merely scaffolding for the development of a moral sense or are constitutive of it.

As I will argue, biological and neurobiological accounts of the origins and development of fine-grained affective responses can no longer be ignored in discussions about the nature of ethics in general and of Buddhist ethics in particular. Of course neuroscience may not be able to tell us, solely on the basis of their neural signatures in the brain, why only certain feelings and dispositions should provide a basis for moral agency. For instance, a disposition to act in a way that shows concern for others, even if habitually acquired, cannot be deemed compassionate if it is not freely undertaken. But such empirical inquiries into the nature of morality *can* tell us whether the kind of moral judgment we associate with compassionate concern for others is primarily driven by affective or cognitive mechanisms.

Does the bodhisattva, the iconic representation of compassionate undertaking, act in a deliberate manner, or is his or her ethical conduct merely

an embodied mode of coping with the situation at hand? Buddhist accounts of moral cultivation agree that for an action to be deemed compassionate it must have been freely undertaken (unlike selfish actions, which are rooted in insurmountable habitual and compulsive tendencies such as greed and delusion). Actions that do not possess the right sort of responsibility-conferring capacity, therefore, fall outside the moral domain. But here we run into a dilemma: insofar as bodhisattvas act compassionately on account of their training and cultivation (presumably they cannot do so otherwise), they can benefit sentient beings habitually or spontaneously (that is, without forming an intention to act in a deliberate manner).⁵ Thus either the bodhisattva's compassionate deeds count as (freely undertaken) moral acts, or they are purely spontaneous, thus mysterious, unpredictable, and outside the framework of moral responsibility. But, of course, nothing can be "outside the framework of moral responsibility"—hence the reference in the title to the television series *Breaking Bad* that chronicles a spectacular example of moral downfall: to "break bad" in this context is to move decisively against the conventions of one's social environment to gain access to a larger nexus of action. By analogy, to "break good" is thus to enter a comprehensive and seemingly unfathomable arena of moral action.⁶

In what follows I will argue that the "spontaneity of compassion" picture we glean from the Mahāyāna ethical literature is problematic in light of the demands placed on our conceptions of moral agency by moral responsibility-entailing practices.⁷ The question, then, is whether the achievement of moral ends requires a more robust conception of agency than the Buddhist no-self view can provide.

6.2. Virtue, Moral Agency, and Consequentialism

One of the most elaborate (and inspiring) accounts of the Buddhist path to moral and mental cultivation is found in Śāntideva's *A Guide to the Path to Awakening* (*Bodhicaryāvatāra*, hereinafter BCA), an immensely influential text that has been interpreted as advancing a version of either consequentialism or virtue ethics. While both ethical theories (and their variants) supply useful conceptual tools for unpacking this (broadly Mahāyāna) Buddhist ethical program, they cannot satisfactorily account for the meta-ethical principles that inform it. My intention is not to showcase the unique features of Buddhist ethics (such as they are) but to ask what Buddhist forms of moral and mental cultivation tell us about the nature of mind and how it can be altered (or, alternatively, about what moral agency is and how it is achieved).⁸ If Buddhist

neuroethics is concerned with the neural basis of enlightened moral agency (specifically the agency of those undergoing secular mindfulness and compassion meditation training), then established correlations between subpersonal processes (specifically those that regulate affective, retributive, and cognitive behavior) and first-person accounts (of what it is like to cultivate and exercise compassion, forbearance, or equanimity) ought to extend rather than limit the scope of moral agency. Indeed any evidence that self-concern may be regulated by mechanisms that also monitor our interest in the well-being of others should provide sufficient ground for advancing a more robust account of the efficacy of compassion meditation practice.

Nietzsche's disquieting attitude toward the ethics of compassion notwithstanding, what explains the appeal of Buddhism in the West is precisely this emphasis on what may be deemed its cardinal virtues: nonviolence, compassion, and a general spirit of tolerance. Of course these virtues are embedded in the theoretical structure of Buddhist ethics (itself part of the Buddhist path writ large)—a project whose characterization lacks scholarly consensus. I will not enter this debate here.⁹ Instead, following recent work in neuroethics and cognitive moral psychology¹⁰—and its implication for analytic reconstructions of Buddhist ethics such as one finds, for instance, in Siderits (2008), Goodman (2009), and Flanagan (2011)—I will ask to what extent the view that moral principles are informed by emotionally driven intuitions rather than, say, deliberate moral reasoning can be said also to apply in the Buddhist context.

Containing what is perhaps the most developed account of moral and mental progress on the bodhisattva path—the iconic representation of a life dedicated to pursuing enlightened knowledge for the sake of benefiting all sentient beings—*A Guide to the Path of Awakening* also showcases the centrality of compassion for Buddhist ethics. Although the cultivation of distinct moral sentiments is suggestive of a virtue-ethical approach, Śāntideva views compassion through what seems like a strongly consequentialist framework. Bodhisattvas with a well-developed character are called upon to exercise judgment when acting in the name of compassion. Thus “for the one who understands the work of compassion even the forbidden is permitted” (BCA 5.84). It is obvious from statements such as these that genuine compassion is accompanied by a level of discretion that permits the bodhisattva to act in peculiar ways; specifically it allows for breaking moral precepts like lying and killing without incurring the retributive effects of these acts (or, at the very least, without the threat of punishment).

That a well-developed character should be enough to mitigate the consequences of (seemingly unfathomable) compassionate acts has led some to

propose that (Mahāyāna) Buddhist ethics is best understood as a type of character consequentialism: the value of generous acts depends on the value they confer upon all those involved, the agent as well as the beneficiary. If compassion is an outcome of character development, and Buddhist flourishing is essentially the embodiment of those perfections deemed essential for the bodhisattva to carry out his or her work in the world, then compassion is not agent-neutral. The generally consequentialist framework of Buddhist ethics, then, cannot be universalist, even though theoretically the bodhisattva is called upon to maximize happiness (and eliminate suffering) for all sentient beings.

As Charles Goodman (2009, p. 43) has quite convincingly argued, agent-neutrality makes consequentialism quite demanding; that is, it calls for great acts of self-sacrifice, which seems not only unrealistic but perhaps unachievable as well. How else is one to interpret Śāntideva's famous aspiration of bringing an end to suffering: "As long as space abides, as long as the world abides, / So long shall I abide, destroying the sufferings of the world" (BCA 10.55)? Whether the ethical ideals of the bodhisattva are suggestive of an agent-neutral framework, and whether that framework makes Buddhist ethics seem closer to consequentialism than to virtue ethics, is precisely what is at stake. For someone like Damien Keown (2001), positive accounts of pleasure and the pursuit of happiness for all sentient beings; such as are found, for instance, in the *Sutra of Golden Light* (*Suvarṇaprabhāsa Sūtra*), suggest that the bodhisattva idea does fit the virtue-ethical model, where the ultimate good is a caring and compassionate love. Developing such obvious virtues as generosity, compassion, and insight is generally how one attains this ultimate good. But these virtues may be also regarded as having an instrumental value, insofar as they tend to counteract various defilements such as greed, hatred, and delusion and to promote progress toward Buddhahood. For Keown these virtues form an intrinsic part of the Buddhist conception of the good. A similar conception of Buddhist flourishing is articulated by Peter Harvey (2000, p. 354), who likewise notes the role certain attitudes and practices can have in fostering the cultivation of such virtues as generosity, compassion, and insight. Of course insofar as the roots of the good here have instrumental value, this ethical model may also be viewed in character-consequentialist terms. Goodman (2009, p. 88) summarizes this view quite well when he writes that "happiness and the absence of suffering, as well as virtues and the absence of vices, are elements on an objective list that defines well-being."

How is this well-being achieved? And what sort of leverage does a bodhisattva (with well-developed character) have in maximizing happiness and minimizing suffering for all beings? It is certainly the case that individuals

not only have different needs but also occupy different rungs of the moral ladder. It is here that the exercise of judgment takes precedence in mediating (or perhaps moderating) the disposition to act in a compassionate way. The Buddhist literature abounds with examples in which enlightened beings use deception to help beginners make progress on the path. (Some of the best examples are those from the *Lotus Sutra*: the Prodigal Son, the Burning House, and the Phantom City.)¹¹ What is the purpose of these examples of deception in the name of a (presumably) higher good? On the one hand, they reflect the specifically Mahāyāna demand that the early teachings of the Buddha be seen as provisory by comparison with such later teaching as one finds in Mahāyāna literature. On the other, they show the cardinal principle of excellence in means (*upāyakauśalya*) at work. For someone like Śāntideva, excellence in means is called for to explain why actions that are proscribed under general ethical precepts (like speaking the truth and causing no harm to others) *could* be permitted. Classic examples are found in such texts as the *Discourse on the Excellence in Means (Upāyakauśalya Sūtra)*, where bodhisattvas are allowed to break standards precepts or rules so long as they are motivated by compassion.¹² On this account, the early teachings do not tell people what is the case, just what is most beneficial to them given their present situation. Such a view would be at odds specifically with Kantian views of morality, since for Kant deception, whatever its ultimate goal, disrespects human dignity. I will return to this issue when considering the Buddhist position on moral responsibility.

Given that Śāntideva does have an explicit position on whether and in what circumstances the interests of some may be weighted against the welfare of others, the expectation is that at least some of the classical features of consequentialism should apply in his case. Goodman identifies, specifically in a passage from the *Compendium on Trainings (Śikṣā-samuccaya*, hereinafter ŚS), not only some but by his count “all” the classical features of act-consequentialism: “the central moral importance of happy and unhappy states of mind; the extension of scope to all beings; the extreme demands; the absence of any room for personal moral space; the balancing of costs and benefits; and the pursuit of maximization” (Goodman, 2009, p. 97). The passage extolls the efforts of the bodhisattva who, “through actions of body, speech, and mind . . . makes a continuous effort to stop all present and future suffering and depression, and to produce present and future happiness and gladness, for all beings” (ŚS, in Goodman, 2009, p. 97).

Given the bodhisattva’s extreme dedication to pursuing the welfare of all sentient beings, the framework of act-consequentialism must provide a way to account for the psychological and neuropsychological mechanisms that could

underpin such ethical conduit. Indeed it is not enough to say that Buddhist ethics is best captured by a specific theory—in this case, consequentialism. One must also ask whether this theory reflects a characteristically Buddhist understanding of the good (e.g., minimizing suffering and maximizing happiness) or simply represents a manifestation of the sort of psychological patterns that are integral to moral sentiments.

6.3. Evidence from Neuroimaging

The Buddhist ethical literature contains frequent references to characteristics like generosity, compassion, and insight that are considered beneficial or wholesome (*kusala*) and thus conducive to achieving the ultimate ends that Buddhists seek. They are typically contrasted with traits like greed, hatred, and delusion, which constitute a major cause of suffering and thus an obstacle to achieving these ends. Both sets of characteristics are classified under the Abhidharma category of mental states (*caitta*). For moral psychologists who wish to get to the roots of morality, the question is this: Which of these mental states are to be understood primarily in affective and which primarily in cognitive terms (even as the classification of mental states in the Abhidharma literature admits of no such distinction)?¹³ It is important to maintain this distinction, given its relevance to neuroimaging studies,¹⁴ which take certain areas of the brain, primarily the dorsolateral surfaces of the prefrontal cortex and parietal lobes, to be associated with cognitive processes, while others, specifically the amygdala and the medial surfaces of the frontal and parietal lobes, track emotional states such as moods and gut feelings.¹⁵

To see whether moral judgments have an emotional component and in what circumstances (and whether) the emotional response may be overridden, let's briefly consider some empirical findings about entertaining the well-known hypothetical moral dilemmas of the trolley and the footbridge.

In the first instance, a runaway trolley is heading for five people gathered some way down a track, who will be killed if the trolley is not diverted onto a sidetrack. The only way to save these people is to divert the trolley. The problem is that there is one person on the alternate sidetrack, who, as a result, will end up being killed. What is one to do? Push the switch so that five people are saved at the expense of one? In the second instance, the same runaway trolley threatens to kill five people, but instead of proximity to a switch, you find yourself on a footbridge next to a large stranger standing right above the track with the runaway trolley. The only way to save the five people is to push this stranger off the bridge and onto the incoming trolley. The stranger will die as

a result, but his body will stop the trolley in its tracks and save the five people. Again, what is one to do?

Neuroimaging studies of subjects presented with these dilemmas show a clear pattern of brain activity. Entertaining the more impersonal moral dilemma of the trolley corresponds to increased activity in brain regions associated with higher cognitive processes like complex planning (see, especially, Koechlin, Basso, Petrini, Panzer, & Grafman, 1999; Koechlin, Ody, & Kouneiher, 2003; Miller & Cohen, 2001) and deductive and inductive reasoning (Goel & Dolan, 2004). On the other hand, the footbridge dilemma activates brain regions associated with strong emotional response (Haidt, 2001; Greene and Haidt, 2002, and Greene et al., 2009). Furthermore engaging in characteristically consequentialist judgments leads to increased activity in those parts of the brain typically associated with higher cognitive functions such as decision making and executive control. The traditional approach to solving such dilemmas typically invokes the normative framework of ethics. The argument is that our response to these dilemmas should be judged relative to norms. Deontologists, for instance, might judge it wrong to harm someone in order to save someone else in all circumstances. The question, then, is not which ethical theory best fits the empirical data but rather what the empirical data tell us about the nature of morality.

As Joshua Greene (2008, p. 43) notes in reviewing the relevant neuroimaging studies, “People tend toward consequentialism in the case in which the emotional response is low and tend toward deontology in the case in which the emotional response is high.” To support his hypothesis, Greene cites evidence from evolutionary history; “up close and personal” violence, for instance, reaches quite far back into our primate lineage by comparison with the types of impersonal harm (e.g., drone strikes) that demand complex forms of abstract reasoning. This hypothesis is further supported by reaction times, as it takes longer to ponder an impersonal moral dilemma (the trolley case) than dilemmas that elicit strong emotional response (the footbridge case). On this account, judging a personal moral violation appropriate (that is, judging that it is permissible to push the stranger to his death in the footbridge case) depends on the capacity to override the emotional response that such an up close and personal action would elicit.

What the evidence from neuroimaging so far suggests is that we are hard-wired to have powerful innate responses to personal violence. We may regard these data as providing evidential support for the view that altruistic behavior is a natural kind. At the same time the capacity to override strong emotional responses implicit in judgments that deem personal moral violations appropriate casts a long shadow on any idea of an innate or intrinsic good.

On the grounds that the cultivation of compassion is essentially a normative aspect of Buddhist practice, we can now ask: Can a bodhisattva on the path of moral and mental cultivation undergo the kind of transformation that renders consequentialist thought immune to emotional response? The Buddhist literature is unambiguous on this point: compassion meditation has quite different effects on beginner bodhisattvas than on those who have advanced along the path. Beginning bodhisattvas are often portrayed as “overwhelmed by compassion” such that they may even be found crying. Progress along the path is such that the occasional outburst has given way to equanimity.¹⁶

Do emotional responses, then, play any role in the generally consequentialist framework of Mahāyāna Buddhist ethics? In the case of prescriptions that sanction moral violations for a greater good, the presumption is that bodhisattvas would have developed the capacity to override their emotional responses, especially those typically associated with inflicting harm on a person judged to be committing some wrongdoing. (The bodhisattva cannot be plagued by an internal struggle to overcome such negative emotions as remorse.) However, such judgments of wrongdoing presuppose a normative framework and imply some notion of moral responsibility. And moral responsibility cannot be understood without addressing the issues of agency and free will.

6.4. Freedom and Human Responsibility

Whether something analogous to the Western notion of free will is found in Buddhism is an open question.¹⁷ A notion of “the will”—in the sense of voluntary action (*voluntas*)—is indeed presupposed by the Buddhist concept of *cetanā*, variously translated as “volition,” “intention,” and “will.” But at the most basic level, *cetanā* captures the dynamic aspects of the mind relative to objects or specific ends. The problem, then, is this: How can accounts of action and its consequences (karma) become part of an ethical framework given the standpoint of Buddhist reductionism?

Are free will and determinism compatible? Is there a way of reconciling our first-person account of volitional action with third-person perspectives of the underlying physical, biological, and now neurobiological processes? Is there another way of conceiving of humans or, indeed, of being human that demands a radical reassessment of our understanding of voluntary action and of the causal and motivational factors that inform, condition, and sanction our valuing judgments? More specifically for our purpose here, how has the relation between volitional and causal accounts of agency been understood in the Buddhist context?

The Buddhist account of this relation originates with Siddhārtha Gautama's experience of enlightenment. This experience becomes at once the source of the Buddhist metaphysical picture of reality and the culmination of all human aspiration for genuine freedom. Key to this metaphysical picture is the causal principle of dependent arising and a thoroughly reductionist account of persons, which takes volition to be but one of the many contributing factors that shape human identity and agency. In one of his earlier discourses, the Buddha declares that we ought to regard any form of sensation, attention, and consciousness, whether "past, future, or present; internal or external; manifest or subtle . . . as it actually is. . . : 'This is not mine. This is not my self. This is not what I am'" (SN.iii.49). Of course the rejection of a permanent self as the agent of sensory, affective, and mental activity poses a significant challenge for Buddhism. If there is no agent, and if actions are merely transient events arising within a continuum of causally interconnected states, how is the intentional orientation of human actions to be understood? Even assuming, as the evidence from cognitive neuroscience seems to suggest, that we are psychologically hardwired to attribute agency and hold others responsible for their actions, the question why such agency-attributing capacities should be accompanied by a moral sense remains to be explained. I will return to this point in my conclusion.

Whether we take the Buddhist no-self view to be a theoretical construct or a descriptive account of the immediacy of lived experience, its picture of human nature and agency undermines (or is irrelevant to) the practical concerns of moral responsibility. Now, as some have suggested (most notably Siderits, 2008), there might be a conflict between the conventional practice of morality (to which the Buddha offers precepts, inspiring tales, and rules of conduct) and Buddhist metaphysical doctrine. The Kantian distinction invoked here, between the concerns of practical reason and theorizing about the nature of things, serves as a useful heuristic: the basic thrust of this broadly Kantian view is that when I engage in theoretical reasoning (of the sort that looks for causal explanation of events) there is no place for concepts like freedom and responsibility. But when I engage in practical reasoning (of the sort that asks "What should I do?" and then looks for the most justifiable course of action) there are good reasons to hold myself responsible for my actions. That is, regardless of whether or not theoretical reason is able to demonstrate freedom, practical reason must assume that freedom is possible for the purpose of action. Bringing this Kantian perspective to bear on the Buddhist account of human agency is motivated by the assumption that the kind of freedom we are supposed to consider (and perhaps criticize) is basically as described by libertarians, that is, as involving complete spontaneity. Furthermore, as the

neuroimaging studies discussed above seem to suggest, the normative framework of deontology may have its roots in a basic tendency to avoid (and thus create circumstances that would minimize) the sort of comportment that is conducive to heightened emotional response.

Do freedom and responsibility, as artifacts of practical reason, belong in a discourse about causation in the natural world?¹⁸ If the concerns of practical and theoretical reason are taken to be mutually entailing, then they do. On the other hand, if theoretical reason is seen to be at odds with our practical concerns about how best to live, then they do not. The Buddhist metaphysical picture of reality, as a product of theoretical reason, is devoid of any reference to selves and their concerns, or indeed to anything substantive. At least in principle, the no-self view would preclude any robust account of freedom and responsibility. Yet Buddhist practice requires the observance of certain norms and the valuation of certain types of thought, speech, and action that are considered beneficial. Chief among these is the restraint of unmitigated willful thought, speech, and action. However, this valuation, and the psychological terms in which it is expressed, is at odds with an impersonal account of the self and subjectivity in causal terms. Siderits's proposal is that some kind of Buddhist compatibilism is called for to solve this conflict. Can such an account, in effect, be offered?

Consider Śāntideva's extension of the permission to break moral rules to those who carry out "compassionate" actions. If such an injunction cannot be easily justified on a normative account of practical reason, then the largely consequentialist framework of compatibilism cannot give an adequate account of our moral institutions. (Such a framework is also indifferent to the concerns of practical reason.) Thus Mahāyāna ethicists are not concerned with the possibility of freedom in a causally ordered universe (such possibility is taken to be the *modus operandi* of all enlightened beings) but with minimizing suffering and/or maximizing happiness. This account of agency in the service of altruistic aims is partly the reason most interpreters have regarded Buddhist ethics as essentially utilitarian or consequentialist in scope.¹⁹ The problem with this account is that it takes moral agency away from ordinary people, who come to be regarded as lacking an understanding of how things are and of the proper motivation for ethical action. (In some sense they are no better than children or the insane.)

Whether conventional morality and the antinomian character of the bodhisattva's conduct are perfectly comprehensible if, as Siderits argues, we take the Buddhist account of practical rationality in straightforwardly consequentialist terms is, indeed, open to debate. Siderits's (2008, p. 39) proposal is that we are dealing here with two versions of consequentialism: indirect consequentialism

for the common folk and act-consequentialism for the bodhisattva, whose enlightened perspective allows for all sorts of shortcuts that are simply not available to the rest of us. But the trust this view places in our capacity to account for motivations that are inscrutable and for responsibilities that are intractable makes the Buddhist view of practical rationality seem rather whimsical.

What, then, do the Buddhists debate when they talk about the Eightfold Noble Path or the cultivation of perfections as an ethical program? And is this seemingly virtue-ethical program compatible with the Buddhist metaphysical picture of reality? It is tempting to say that what is at stake here is the so-called mechanics of salvation, that is, whether or not the disciplined cultivation at the heart of the Eightfold Noble Path guarantees the liberation from suffering and cyclical existence that Buddhists aspire to. The question is this: Can disciplined cultivation take the place of practical reason?

Certainly the fact that ethical concerns occupy only the lower rung of the Eightfold Path program would suggest that for the Buddhist moral norms are conventional and ultimately they should be overcome or even discarded. What does that mean for our understanding of the relation between freedom and responsibility? It is hard to say.²⁰ The idea that there are types of freedom, specifically freedom from suffering and rebirth, that are not responsibility-entailing (at least not in terms of reasons for which actions might be held accountable) seems to advocate a type of libertarian agency that is hard to reconcile with Buddhist reductionism.

6.5. Conclusion

When Śāntideva allows for moral rules to be violated under the expediency of a compassionate aim, he likewise undermines the traditional notion of responsibility.²¹ Of course the absence of a strictly causal account of action poses an even greater threat than does any notion of determinism. If prior conditioning does not determine our thoughts and actions, then they must be either random or spontaneous. And indeterminism does not make things any easier for the compatibilist than determinism does; quite the contrary (bumper-sticker wisdom of the sort that urges us to “practice random acts of kindness” notwithstanding). Neocompatibilist positions such as one finds, for instance, in Flanagan (2002), address some of these challenges by showing how, if we dispense with the incoherent notion of libertarian agency, some notion of responsibility can be salvaged. For Flanagan and all neocompatibilists who recognize the need to take cognitive science seriously, the main issue is that our traditional notions of agency and responsibility are in need of revision. Indeed recent advances in the study of human cognition suggest that much

of our conscious mental life depends on subconscious cognitive processes. In setting out to offer an account of how the two pictures are compatible, neocompatibilists shift the dependency relation for freedom and responsibility from norms to facts. Critics of this sort of approach, and here I would include myself, point out that the results of cognitive science are subject to constant revision and that a notion of moral agency explainable in terms of, say, dispositions is provisory at best.²² Indeed the normative features of moral reasoning do not sit well with the revisionist methods of science, at least not as traditionally understood, even though there are no good reasons to exclude moral norms from the purview of empirical research.

Now just as conscious awareness remains the single most puzzling and most difficult phenomenon to explain in reductive terms, the patterns exhibited by moral agency also resist the eliminative reductionism of certain types of scientific explanation. Some philosophers of cognitive science argue that human consciousness is inherently intersubjective, and therefore empathy must count as a precondition of consciousness (Thompson, 2001). This line of argumentation suggests that agency presupposes some degree of self-awareness and of concern for others, both of which resist impersonal causal explanation. Maybe the Buddhist ethicists have in mind a similar sort of resistance when they allow for the compassionate aspirations of the bodhisattva to trump psychological determinism. However, if the bodhisattva can attain a type of freedom that is unimpeded by karmic hindrances, the efficacy of his or her actions (outside the web of interdependent causation) becomes deeply mysterious. After all, as we noted above, tales of bodhisattvas who intervene, as if magically, to take humans out of the trap of cyclical existence abound in the Buddhist literature. Does that make the bodhisattva a sort of compassionate libertine? Perhaps. But in that case genuine compassion implies a kind of spontaneity that is not easily captured by notions of moral agency that depend only on the actual or foreseeable consequences of acts. Whether such compassionate spontaneity also possesses the right sort of responsibility-conferring capacity cannot be settled without further probing the mesh that is the bodhisattva's practical wisdom and skillful concern for others.

Abbreviations

- BCA Śāntideva's *Bodhicaryāvatāra*. Translations from K. Crosby & A. Skilton in Śāntideva (1995).
- SN *Samyutta Nikāya*, volume and page in the Pali Text Society edition. Translations from Bodhi (2000).
- SS Śāntideva's *Śīkṣā-samuccaya*. In Śāntideva (2016).

Notes

1. The exception here is Nussbaum's (2001) comprehensive account of the cognitive structure of compassion in the context of an analysis of what she regards, borrowing a concept from Greek theater, as the tragic predicament of the human condition. Taking as her point of departure Aristotle's view of compassion—which, following Homer and Plato, he regards as a painful emotion directed at another person's misfortune—Nussbaum aligns herself with those who defend the centrality of compassion to any moral theory (e.g., Rousseau, Schopenhauer, and Adam Smith) against the opponents of emotion (e.g., the Greek and Roman Stoics, Spinoza, Kant, and Nietzsche). Noting the centrality of altruistic concerns to certain philosophical conceptions of morality (especially those that appeal to evolutionary theory to make the case that psychological altruism is true), Stich, Doris, and Roedder (2010) rightly identify compassion as the sort of "right emotion" for producing moral motivation. Nonetheless, they pay only scant attention to the role of compassion in explaining the link between moral motivation and voluntary action.
2. An extensive review of the literature on affective neuroscience and compassion is found in Davidson (2002). See also Lamm, Batson, and Decety (2007), Lutz, Brefczynski-Lewis, Johnstone, and Davidson (2008), and Mascaro, Rilling, Tenzin Negi, and Raison (2012) for studies that showcase high empathic responses in individuals who undergo a form of secularized analytical compassion meditation.
3. See Coseru (2012, Chapter 3.3). Note that "intentional content" here is broadly conceived to include both what the mental state is a state of (its object) as well as its own operations. Of course whether intentional content is distinct from the objects intended by such mental acts as perceiving or judging is a controversial topic. For more on this debate, see Zahavi (2004).
4. Wilson's (1998, p. 54) suggestion that "causal explanation of brain activity and evolution" suffices as an explanation of moral behavior may be unwarranted in light of the fact that moral judgment has the capacity to alter conditioned behavior in substantive ways; being made aware, for instance, that one is evolutionarily conditioned to favor members of one's group over outsiders can, on reflection, lead to adopting a more egalitarian view. Although tracing the history of moral behavior is notoriously difficult, its roots are generally assumed to lie in an innate biological altruism (which is found in many species). It is this biological altruism that underpins the more complex psychological altruism at the heart of our so-called ethical project (cf. Kitcher 2011). Indeed only the latter can explain how we come to care for the welfare of others not for our but for their own sake.
5. For an argument in favor of the spontaneity of compassion, see Williams (2009, pp. 116–117). In his discussion of BCA 9.34–35, Williams takes the view that budhas are disposed to help precisely because they lack reifying and modal propositional attitudes.

6. I owe this suggestion to Sheridan Hough, who explores the causal aspects of agency and intersubjectivity in her essay on *Breaking Bad*; as she points out (Hough, 2016, p. 218), the series creator, Vince Gilligan, “denies the discrete or independent reality of the objects and persons within the causal structure.” As Gilligan himself puts it: “I like to believe . . . that karma kicks in at some point, even if it takes years or decades to happen” (quoted in Hough, 2016, p. 218).
7. I explore this issue at length in Coseru (2016), where I also address the tension between the Buddhist account of the irreducibility of the mental and the tendency to confine freedom and responsibility to the domain of social convention. If the Buddhist’s ultimate ontology contains phenomenal primitives, then freedom and responsibility cannot be mere artifacts of conceptual proliferation.
8. While the secondary literature of Buddhist ethics has grown considerably in recent years, with few exceptions (notably Keown, 1996; Siderits, 2008; Goodman, 2009; Garfield, 2010; Finnigan, 2011), most treatments are still exegetical in scope and anchored in specific texts and/or traditions. Surveys that address the relevance of Buddhist ethical principles to a wider range of topics, including such controversial issues as human rights, animal rights, ecology, war, and abortion, are found in Cooper and James (2005) and Keown (2007).
9. Readers may consult the excellent surveys in Harvey (2000), Keown (2001), Clayton (2006), and Goodman (2009). I address some of the same issues in Coseru (2016).
10. See especially Farah (2005), Gazzaniga (2005), Greene and Haidt (2002), and Greene (2008, 2009).
11. For a close look at the expedient role of these parables, with specific reference to the Prodigal Son, in the *Lotus Sutra*, see Lai (1981).
12. An ethics of compassionate violence, as Stephen Jenkins (2010, p. 326) has convincingly argued, works by “removing the possibility that any action is essentially inauspicious.” Such an ethics, then, coupled with the (metaphysical) notion that the workings of karma are generally inconceivable, not only engenders ambiguity about the bodhisattva’s ethical program but also diminishes the capacity for moral certainty.
13. See Dreyfus (2002) for an illuminating account of the difficulties of cross-cultural approaches to mental typologies with specific reference to the dialogue between Buddhism and cognitive science.
14. See, especially, Koechlin, Ody, and Kouneiher (2003), Miller and Cohen (2001), and Ramnani and Owen (2004).
15. See, especially, Adolphs (2002), Maddock (1999), and Phan, Wager, Taylor, and Liberzon, (2002).
16. Contrast, for example, BCA 2.50, “In despair, I cry out for help to protector Avalokiteśvara, who acts compassionately and inerrantly, begging him to protect my vicious self”, with BCA 10.2, “Whoever is suffering distress of body or

mind in any of the ten directions—may they obtain oceans of happiness and joy through my good actions.”

17. See Dasti and Bryant (2014) and Repetti (2016), and contributions therein.
18. Addressing a similar issue, but with respect to the moral implications of advances in neuroscience, Hilary Bok (2007) concludes that learning about the many ways freedom can be undermined (by phobias, compulsions, failures of self-control, etc.) offers an opportunity to conceive of freedom in more effective terms. Bok proposes that we understand freedom as a “capacity for self-governance” rather than as a type of choice-driven action (p. 559).
19. Proposals that advance a utilitarian interpretation of Buddhist ethics are found in, among others, Pratt (1928), Kalupahana (1976), and Goodman (2009).
20. I do, however, venture a response to this question in Coseru (2016).
21. See BCA 5.83 and ŚS 168.11, and discussion in Clayton (2009, p. 23).
22. Stich, Doris, and Roedder (2010) reach the same conclusion about employing evolutionary theory to intervene in the empirical debate between altruists and egoists: useful as it may be, in the end, neither camp makes a convincing case.

References

- Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, 12(2), 169–177.
- Bodhi, B. (Trans.). (2000). *The connected discourses of the Buddha*. Boston: Wisdom.
- Bok, H. (2007). The implications of advances in neuroscience for freedom of the will. *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics*, 4, 555–559.
- Brefczynski-Lewis, J. A., Lutz, A., Schaefer, H. S., Levinson, D. B., & Davidson, R. J. (2007). Neural correlates of attentional expertise in long-term meditation practitioners. *Proceedings of the National Academy of Sciences*, 104, 11483–11488.
- Brooks, D. (2008, May 13). The neural Buddhists. *New York Times*.
- Clayton, B. R. (2006). *Moral theory in Śāntideva’s Śīksāsamuccaya: Cultivating the fruits of virtue*. New York: Routledge.
- Clayton, B. R. (2009). Śāntideva, virtue, consequentialism. In J. Powers & C. S. Prebish (Eds.), *Destroying Mara forever* (pp. 16–29). New York: Snow Lion.
- Cooper, D. E., & James, S. P. (2005). *Buddhism, virtue, environment*. Burlington, VT: Ashgate.
- Coseru, C. (2012). *Perceiving reality: Consciousness, intentionality, and cognition in Buddhist philosophy*. New York: Oxford University Press.
- Coseru, C. (2016). Freedom from responsibility: Agent neutral consequentialism and the bodhisattva ideal. In R. Repetti (Ed.), *Buddhist perspectives on free will* (pp. 92–105). London: Routledge.

- Dasti, M., & Bryant, E. F. (Eds.). (2014). *Free will, agency, and selfhood in Indian philosophy*. New York: Oxford University Press.
- Davidson, R. (2002). Toward a biology of positive affect and compassion. In R. Davidson & A. Harrington (Eds.), *Visions of compassion: Western scientists and Tibetan Buddhists examine human nature* (pp. 107–130). New York: Oxford University Press.
- Dreyfus, G. (2002). Is compassion an emotion? A cross-cultural exploration of mental typologies. In R. Davidson & A. Harrington (Eds.), *Visions of compassion: Western scientists and Tibetan Buddhists examine human nature* (pp. 31–45). New York: Oxford University Press.
- Farah, M. J. (2005). Neuroethics: The practical and the philosophical. *Trends in Cognitive Sciences*, 9, 34–40.
- Finnigan, B. (2011). How can a Buddha come to act? The possibility of a Buddhist account of ethical agency. *Philosophy East and West*, 61(1), 134–160.
- Flanagan, O. (2002). *The problem of the soul: Two visions of the mind and how to reconcile them*. New York: Basic Books.
- Flanagan, O. (2011). *The bodhisattva's brain: Buddhism naturalized*. Cambridge, MA: MIT Press.
- Garfield, J. (2010). What it is like to be a bodhisattva? Moral phenomenology in Śāntideva's *Bodhicaryāvarāta*. *Journal of the International Association of Buddhist Studies*, 33(1–2), 333–358.
- Gazzaniga, M. S. (2005). *The ethical brain: The science of our moral dilemmas*. New York: Dana Press.
- Goel, V., & Dolan, R. J. (2004). Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition*, 93, B109–B121.
- Goodman, C. (2009). *Consequences of compassion: An interpretation and defense of Buddhist ethics*. New York: Oxford University Press.
- Greene, J. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology* (vol. 3, pp. 35–79). Boston: MIT Press.
- Greene, J. (2009). The cognitive neuroscience of moral judgment. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences IV*. Cambridge, MA: MIT Press.
- Greene, J., Cushman, F. A., Stewart, L.A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, 111, 364–371.
- Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, 6(12), 517–523.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.
- Harvey, P. (2000). *An introduction to Buddhist ethics*. Cambridge, UK: Cambridge University Press.
- Hough, S. (2016). “We are responsible to all for all”: an intersubjective analysis of *Breaking Bad*. In Decker, K. S., Koepsell, D. R., & Arp, R., (Eds.), *Philosophy and Breaking Bad* (pp. 217–232). Palgrave Macmillan.

- Jenkins, S. (2010). On the auspiciousness of compassionate violence. *Journal of the International Association of Buddhist Studies*, 33(1–2), 299–332.
- Kalupahana, D. (1976). *Buddhist philosophy: A historical analysis*. Honolulu: University Press of Hawaii.
- Kant, I. (1993). *Grounding for the metaphysics of morals*. (J. W. Ellington, Trans.). Indianapolis: Hackett.
- Keown, D. (1996). Karma, character, and consequentialism. *Journal of Religious Ethics*, 24(2), 329–350.
- Keown, D. (2001). *The nature of Buddhist ethics*. London: Macmillan.
- Keown, D. (2007). Buddhism and ecology: A virtue ethics approach. *Contemporary Buddhism*, 8(2), 97–112.
- Kitcher, P. (2011). *The ethical project*. Cambridge, MA: Harvard University Press.
- Koechlin, E., Basso, G., Petrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature*, 399, 148–151.
- Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, 302, 1181–1185.
- Lai, W. (1981). The Buddhist “prodigal son”: A history of misperceptions. *Journal of the International Association of Buddhist Studies*, 4(2), 91–98.
- Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience*, 19(1), 42–58.
- Lutz, A., Brefczynski-Lewis, J., Johnstone, T., & Davidson, R. J. (2008). Regulation of the neural circuitry of emotion by compassion meditation: Effects of meditative expertise. *PLoS ONE*, 3(3), e1897.
- Lutz, A., Dunne, J., & Davidson, R. (2007). Meditation and the neuroscience of cognition. In P. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 499–554). Cambridge, UK: Cambridge University Press.
- Maddock, R. J. (1999). The retrosplenial cortex and emotion: New insights from functional neuroimaging of the human brain. *Trends in Neuroscience*, 22, 310–316.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.
- Mascaro, J. S., Rilling, J., Tenzin Negi, L., & Raison, C. L. (2012). Compassion meditation enhances empathic accuracy and related neural activity. *Social Cognitive & Affective Neuroscience*, 5, 1–8.
- Nietzsche, F. (2006). *On the genealogy of morality*. (C. Diethe, Trans.). Cambridge, UK: Cambridge University Press.
- Nussbaum, M. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge, UK: Cambridge University Press.
- Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*, 16, 331–348.
- Pratt, J. B. (1928). *The pilgrimage of Buddhism*. New York: Macmillan.

- Ramnani, N., & Owen, A.M., (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nat Rev Neurosci*, 5, 184–194.
- Repetti, R. (Ed.). (2016). *Buddhist perspectives on free will: Agentless agency?* London: Routledge.
- Stich, S., Doris, J. M., & Roedder, E. 2010. Altruism. In John M. Doris & The Moral Psychology Research Group (Eds.), *The moral psychology handbook* (pp. 147–205). Oxford: Oxford University Press.
- Śāntideva. (1995). *The Bodhicaryāvatāra*. (K. Crosby & A. Skilton, Trans.). Oxford: Oxford University Press.
- Śāntideva. (2016). *The Training Anthology of Śāntideva: A Translation of Śikṣā-samuccaya*. (C. Goodman, Trans.). New York: Oxford University Press.
- Siderits, M. (2008). Paleo-compatibilism and Buddhist reductionism. *Sophia*, 47, 29–42.
- Thompson, E. (2001). Empathy and consciousness. *Journal of Consciousness Studies*, 8(5–7), 1–32.
- Williams, P. (2009). Is Buddhist ethics virtue ethics? In J. Powers & C. S. Prebish (Eds.), *Destroying Mara forever* (pp. 114–137). New York: Snow Lion.
- Wilson, E. O. (1998, April). The biological basis of Morality. *Atlantic Monthly*, 53–70.
- Zahavi, D. (2004). Husserl's noema and the internalism-externalism debate. *Inquiry*, 47(1), 42–66.