Consequentialism, Compassion, and Buddhist Neuroethics

Christian Coseru


I. Introduction:

Although the credit for announcing the advent of “Neural Buddhism” perhaps goes to op-ed New York Times 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 elloquent (and disquietingly) proclaimed its arrival more than a century ago. In On the Genealogy of Morality, Nietzsche speaks of the morality of compassion, and of the many (by his own estimation sinister) ways in which it has cast around “even wider to catch even philosophers and make them ill”\(^1\)—the philosophers’ sickness being nothing but the symptom of a culture about to give birth to a new “European Buddhism.”

The contemporary metaethical literature broadly considered lacks a moral account of compassion. 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, a subfield of Neurobuddhism. 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 Patricia Churchland’s ‘neurophilosophy’ and Francisco Varela’s ‘neurophenomenology,’ terms coined to

\(^1\) Nietzsche (2006: 7).

\(^2\) An extensive review of the literature on affective neuroscience and compassion is found in
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, ‘neurobuddhism’ then describes a 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.

This paper addresses two specific questions the Buddhist neuroethics program raises for our traditional understanding of Buddhist ethics: first, does affective neuroscience supply enough evidence for a naturalized account of compassion? Second, 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

² An extensive review of the literature on affective neuroscience and compassion is found in Davidson (2002), See also Mascaro et al. (2012) for a recent study showcasing high empathic responses in individuals who undergo a form of secularized analytical compassion meditation.
compatibilism. Specifically, neocompatibilism is regarded as capable of accommodating freedom and determinism because it sees 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 above, I want to argue that such dispositions 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, it also reflects the generally naturalist outlook of Buddhist moral psychology: indeed, given that dispositions and reasons have an event structure (they are constituted as mental states with specific intentional content) they are also causes, or at least are causally relevant for action. 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 forms judgments. If that is the case, then the roots of morality lie much deeper in the structure of (conscious) behavior than one might think.

\[\text{\textsuperscript{3} See Coseru (2012: 301 and chapter 3.3). Note that ‘intentional content’ here is broadly conceived to include both what the mental state is a state of (it’s objects) as well as its own operations. Of course, whether intentional content is distinct from the objects intended by such mental acts as perceiving or judging is a controversial topic. For more on this debate, see Zahavi (2004).}\]
In response to the second question, I argue 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, at least some of its features should be easily accommodated by the dynamic of social and interpersonal relations.

Nonetheless, even as a late achievement moral responsibility 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.\(^4\) At least for now the jury is still out on whether genetic, environmental, social, and interpersonal factors provide merely scaffolding for the development of morality or are constitutive of it.

\(^4\) See Lutz, Dunne, & Davidson (2006) and Brefczynski-Lewis et al. (2007).
As I will argue in the next section, 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. True, neuroscience cannot tell us why only certain feelings and dispositions should provide a basis for moral agency. 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 it 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 response merely an embodied mode of coping with the situation at hand? At least on the traditional Buddhist account, for an action to be deemed compassionate it must have been freely undertaken (the same presumably is not true of selfish actions since these are rooted in insurmountable habitual and compulsive tendencies, like greed). That is, actions that do not possess the right sort of responsibility-conferring capacity fall outside the moral domain. Insofar as bodhisattvas act compassionately (presumably they cannot do so otherwise), they can actually benefit sentient beings spontaneously without forming an intention to act in a deliberative manner.\(^5\) I think the ‘spontaneity of compassion’ picture is deeply flawed, but first we need to clarify what the achievement of moral agency consists in to understand why might that be the case.

---

\(^5\) For an argument in favor of the spontaneity of compassion, see Williams (2009: 116f). In his discussion of BCA 9: 34-35, Williams takes the view that Buddhas are disposed to help precisely because they lack reifying and modal propositional attitudes.
II. 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), a immensely influential text that has been interpreted as advancing a version of either consequentialism or virtue ethics. I want to argue that 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 metaethical 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). If Buddhist neuroethics is concerned with the neural basis of moral agency (specifically the agency of those undergoing (or having undergone) secular mindfulness and compassion meditation training), then established correlations between subpersonal processes (specifically, those that regulate affective, retributive, and cognitive behavior) and first-personal accounts (of what it is like to cultivate and exercise compassion, forbearance or

---

6 While the secondary literature of Buddhist ethics has grown considerably in recent years, with few exceptions (notably Keown (1996), Siderits (2008) Goodman (2009), and 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, abortion, and so on, are found in Cooper and James (2005) and Keown (2007)
equanimity) ought to extend rather than limit the scope of those practical deliberations that define the character 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.

Nietzsche’s disquieting attitude toward the ethics of compassion notwithstanding, what many Western philosophers find attractive in Buddhism is precisely its emphasis on what may be deemed its cardinal virtues: nonviolence, compassion, and a general spirit of tolerance. But these virtues are embedded in the theoretical structure of Buddhist ethics (itself part of the Buddhist path writ large) and, as is well known, agreement about how best to characterize this ethics is yet to be reached. I will not enter this debate here. Instead, following recent work in neuroethics and cognitive moral psychology and its implication for analytic reconstructions of Buddhist ethics such as one finds 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 deliberate moral reasoning can be said also to apply in the Buddhist context.

Containing what is perhaps the most developed account of the progression on the bodhisattva path, the iconic representation of a life dedicated to pursing enlightened knowledge for the sake of benefiting all sentient beings, A Guide to the Path of

---

7 Readers may consult the excellent surveys in Harvey (2000), Keown (2001), and Goodman (2009).

8 See especially, Farah (2005), Gazzaniga (2005), Greene and Haidt (2002), and Greene (2008, 2009).
Awakening also showcases the centrality of compassion for Buddhist ethics. But 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 like this one 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 (at times unfathomable) compassionate acts has led some to propose that 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 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 as doctrinally the bodhisattva is called upon to maximize happiness (and eliminate suffering) for all sentient beings.

Charles Goodman (2009: 43) has recently argued that agent-neutrality makes consequentialism quite demanding: that is, it calls for great acts of self-sacrifice, which seem 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 are found, for instance, in the Sutra of Golden Light, suggest that the bodhisattva idea does fit the virtue ethical model where the ultimate good is a carrying 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 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 (Harvey 2000: 354), who likewise notes the role that 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 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” (Goodman 2009: 88).

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 the spiritually undeveloped 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\(^9\)). 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 demands that the early teachings of the Buddha be seen as provisory by comparison with the latter teaching found 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 texts like the Discourse on the Excellence in Means (Upāyakauśalya Sūtra), where bodhisattvas are allowed to break standards precepts or rules so long as they act out of compassion.\(^{10}\) On this account, the early teaching do not tell people what is the case, just what is most beneficial to them given

---

\(^9\) 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).

\(^{10}\) An ethics of compassionate violence, as Stephen Jenkins has convincingly argued, works by “removing the possibility that any action is essentially inauspicious” (Jenkins 2010: 326). But then such an ethics, 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.
their present situation. Such a view would be at odds specifically with the Kantian view of morality, since for Kant deception, whatever its ultimate goal, disrespects human dignity. I will return to this issue below when considering the Buddhist position on moral responsibility.

Since 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* (Śīkṣā-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: 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.”

The framework of act-consequentialism with its extreme dedication to pursuing the welfare of all sentient being demands some explanation of the psychological and neuropsychological underpinnings of this ethical conduit. Indeed, it is not enough to say that the framework of Buddhist ethics is best captured by a specific theory, in this case, consequentialism. One must also ask whether this specific theory reflects a certain understanding of the good as conceived in the Buddhist context (in this case, minimizing

---

11 Quoted in Goodman (2009: 97).
suffering and maximizing happiness) or can be regarded simply as a manifestation of psychological patterns that are typically associated with different moral sentiments.

**III. Evidence from Neuroimaging**

The Buddhist ethical literature contains frequent references to characteristics like generosity, compassion, and insight that are considered beneficial or wholesome (*kuśala*), 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. But both sets of characteristic are classified under the Abhidharma category of mental states (*caitta*). For the naturalist who wants to get to the roots of morality, the question is 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\(^\text{12}\)). This distinction is important since from the perspective of neuroimaging studies\(^\text{13}\) certain areas of the brain, primarily the dorsolateral surfaces of the prefrontal cortex and parietal lobes are associated with cognitive processes, whiles others, specifically the amygdala and the medial surfaces of the frontal and parietal lobes are

---

\(^{12}\) 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.

\(^{13}\) See, especially, Koechlin, Ody, & Kouneiher (2003), Miller & Cohen (2001), and Ramnani & Owen (2004).
associated with emotional states such as moods and gut feelings.\textsuperscript{14}

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 some way down a track, which will be killed if the trolley is not diverted onto a sidetrack. The only way to save these people is to divert the trolley such that it will only end up killing one person instead of five. 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 close to a switch you find yourself on a footbridge next to a large stranger standing right above the track of 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,\textsuperscript{15} and deductive and inductive reasoning (Goel & Dolan, 2004). On the other hand, the footbridge dilemma activates brain regions associated with


\textsuperscript{15} See, especially, Koechlin et al. (1999, 2003) and Miller & Cohen (2001).
strong emotional response (Greene et al., 2004). 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 responses to these dilemmas should be judged relative to norms. Following Kant, one might judge it wrong to harm someone in order to save someone else in all circumstances. The question here is not which ethical theory best fits the empirical data, but rather what the empirical data tells us about the nature of morality.

As Joshua Greene 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” (Greene 2007: 42). 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 hardwired to have powerful innate responses to personal violence. We may regard this data as some sort of evidential proof 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 to be appropriate casts a long shadow on any idea of an innate or intrinsic good.

Since the possibility of cultivating 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 clear that the cultivation of compassion has quite different effects on beginner bodhisattvas than on those who have advanced along the path. Beginning bodhisattvas are often portrayed as being “overwhelmed by compassion” such that they may be even found crying. Progress along the path ensures that compassion as it becomes stronger also achieves balance as the occasional outbursts give way to equanimity.\(^\text{16}\)

Do emotional responses, then, play any role in the generally consequentialist framework of Mahāyāna Buddhist ethics? In the case of prescriptions in the Buddhist literature where a moral violation is sanctioned for greater good, the presumption is that the bodhisattva through repeated practice has developed the capacity to override

\(^{16}\) 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”).
emotional responses typically associated with inflicting harm on a person judged to be committing some wrongdoing. But such judgments of wrongdoing presuppose a normative framework and imply some notion of moral responsibility. And moral responsibility cannot be understood without addressing the issue of free will.

IV. Freedom and Human Responsibility

Whether something analogous to the Western notion of free will is found in Buddhist traditions is debatable and interpretations range widely. Regardless, addressing this last issue is essential if we are to understand the character of moral agency in Buddhism. Western views of the relationship between agency and determinism began to take shape at the dawn of the modern era, marked in part by a revival of interest in classical moral theories and their preoccupation with virtue and human flourishing. They also have their roots in a critique of the rigid and presumably infallible principles of Judeo-Christian morality advanced by French moralists philosophers like Montaigne, Pascal and La Rochefoucauld, and in the emergence of a new type of human that no longer occupies a fixed place is the natural order (as imagined by Renaissance thinkers like Pico della Mirandola). It is this free-willing human, with all his or her foibles, dispositions, and candid aspirations of freedom that becomes the harbinger of enlightenment era conceptions of rationality and of all subsequent Western disputes about whether morality is (or should be) grounded in the laws of nature or in the laws of reason.

Indeed, a great deal of modern thought is concerned with individual autonomy and the unlocking of human potential in a world that is not longer seen as governed by inscrutable divine laws. While the existence of free will is not always implicitly assumed (except perhaps by anarchists and extreme libertarians), the main concern seems to be
with how best to express it. Those who question the possibility of free will in a deterministic universe need look no further than to the great personal and social upheavals of our recorded history. Critics, arguing from a naturalist standpoint (and drawing extensively on social, biological, and neurobiological research on free will), object that what is here invoked as an expression of free will is but the messy unraveling of the evolutionary drama of the human animal, constrained by its phylogeny and bound by environmental factors to a course of action that is ultimately reducible to natural causes: climate change, disease, the scarcity or abundance of resources, genetic mutations, and other impersonal factors.

Are free will and determinism compatible? Is there a way of reconciling our first-personal account of volitional action with third-person perspectives of the underlying physical, biological, and now neurobiological processes? Is there another, perhaps more enlightened 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 human 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 psychological 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’” (Samyutta Nikāya 22, 48 in Bodhi (2000: 887)). Of course, the rejection of a permanent self as the agent of sensory, affective, and mental activity poses a significant challenge for the Buddhist. Indeed, if there is no agent, and if actions are merely transient events arising within a continuum of causally interconnected states, it becomes difficult to explain the intentional orientation of human actions. 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, I claim, 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 an useful heuristic: the basic thrust of this broadly Kantian approach 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 theoretical reason is able to demonstrate freedom or not, 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 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 inductive of heightened emotional response.

Do freedom and responsibility, as artifacts of practical reason, belong in a discourse about causation in the natural world?17 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

17 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 in which 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 (Bok 2007: 559).
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 phenomena in causal terms. Siderits’ proposal is that some kind of Buddhist compatibilism is called for to solve this conflict. Can such an account, in effect, be offered? My view is that it cannot, for the largely consequentialist framework of compatibilism cannot give an adequate account of our moral institutions, and is generally indifferent to the concerns of practical reason.

Indeed, for Mahāyāna ethicists like Śāntideva, the permission to break a moral rule is extended to those who carry out ‘compassionate’ actions, something that is not easily justified on a normative account 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 why most interpreters have regarded Buddhist ethics as essentially utilitarian or consequentialist in scope.\textsuperscript{18} 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 compassionate action (they are seen in some sense as 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

\textsuperscript{18} Proposals that advance a utilitarian interpretation of Buddhist ethics are found in, among others, Pratt (1928), Kalupahana (1976), and Goodman (2009).
of practical rationality in straightforwardly consequentialist terms is, however, open to debate. Siderits’ 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 (Siderits, 2008: 39). But the trust this view places in the 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 broadly virtue ethical program, and its strictly deterministic account of moral psychology, 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. My question is: can disciplined cultivation thus 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 should ultimately be overcome. What does that do for our understanding of the relation between freedom and responsibility? It is hard to say. 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 the strict determinism of Buddhist psychology.

V. 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. 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 random. 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 neo-compatibilists 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. Likewise, as we noted above, the growing field of experimental philosophy has begun to call into question much of our traditional understanding of the role that moral intuitions play in any theory thereof. 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 I include myself among them, 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. But morality does 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 like Evan Thompson (2001) argue that human consciousness is inherently intersubjective, and therefore empathy must count as a precondition of consciousness. 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. But if the bodhisattva can attain a type of freedom that is unimpeded by karmic hindrances, the efficacy of her actions outside the web of interdependent causation is 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 answered without further probing the mesh that is the bodhisattva’s practical wisdom and skillful concern for others.

Bibliography


