Journal of Buddhist Ethics
ISSN 1076-9005
http://www.buddhistethics.org/
Volume 18, 2011

Paternalist Deception in the *Lotus Sūtra*: A Normative Assessment

Charles A. Goodman
Binghamton University

Paternalist Deception in the *Lotus Sūtra*: A Normative Assessment

Charles A. Goodman[1]

Abstract

The *Lotus Sūtra* repeatedly asserts the moral permissibility, in certain circumstances, of deceiving others for their own benefit. The examples it uses to illustrate this view have the features of weak paternalism, but the real-world applications it endorses would today be considered strong paternalism. We can explain this puzzling feature of the text by noting that according to Mahāyāna Buddhists, normal, ordinary people are so irrational that they are relevantly similar to the insane. Kant's determined anti-paternalism, by contrast, relies on an obligation to see others as rational, which can be read in several ways. Recent work in psychology provides support for the *Lotus Sūtra's* philosophical anthropology while undermining the plausibility of Kant's version. But this result does not necessarily lead to an endorsement of political paternalism, since politicians are not qualified to

[1]Department of Philosophy, Binghamton University, Binghamton, NY 13902. Email: cgoodman@binghamton.edu .

wield such power. Some spiritual teachers, however, may be morally permitted to benefit their students by deceiving them.

Autonomy, Paternalism, and Appropriate Means

The *Lotus Sūtra* has had a profound influence on the lives of hundreds of millions of people, mainly in East Asia. Although many have turned to it for advice about how to live, and although it does contain ethical guidance, the *Lotus Sūtra* is not primarily a work of ethics. But we do find in the text, repeated over and over again, a particular kind of moral view; a view, in fact, that is potentially both troubling and controversial. Numerous passages in the *Lotus Sūtra* present us with wise, compassionate figures who assist others by lying to them, or at least, making misleading statements to them. Some ethicists in the Western tradition, most notably Immanuel Kant, would deny the moral permissibility of this way of helping others. Could spiritual teachers such as the Buddha be justified in using deception to promote the happiness and spiritual growth of their students? Or is there a moral error at the heart of the *Lotus Sūtra*?

There is already a rich tradition of scholarly inquiry into the *Lotus Sūtra*. Therefore, in investigating the question of paternalist deception in that text, I do not need to start from scratch. Damien Keown and Gene Reeves, among other scholars, have already explored these issues; I shall summarize a few of their conclusions. The benevolent lies discussed in the *Lotus Sūtra* are classified under the important Buddhist concept of *upāya*. The most common translation of this term is "skillful means," but Gene Reeves suggests it should be rendered as "appropriate means" (*Appropriate* 382). Doctrinally, the concept of appropriate means is repeatedly used to account for the existence of the Way of the Disciples (Skt. *śrāvaka-yāna*) as one authentic form of the Buddha's religion, and to explain the differences between that teaching and the Mahāyāna.

From a Western point of view, the benevolent lies in question would evidently count as paternalism. This is deceptive, rather than coercive, paternalism: the *Lotus Sūtra* never explicitly endorses paternalist uses of physical force, and at one point seems to reject them.[2] Moreover, as Keown points out, in most of the parables, the characters in the stories who benefit from being deceived are not fully rational: they are children, or mentally disturbed, or tired and distressed, and so on (373). So if we focus on these stories, what the *Lotus Sūtra* endorses would count as weak paternalism, which is the use of coercion or deception to get people who are for whatever reason not fully rational to do or allow what would be in their best interests. Weak paternalism is not nearly as controversial as strong paternalism, in which coercive or deceptive means are employed on normal, adult humans whose rationality is not impaired by any unusual conditions. But if we bring our attention to the doctrinal applications of the parables in question, the people who are being deceived are normal adults, including many of the Buddha's own Disciples. So although the parables themselves are examples of weak paternalism, they are being used to support a view that, by ordinary standards, appears to count as strong paternalism.

Having implicitly recognized this point, Keown decides not to wrestle with the issues it raises, writing:

> How accurately these parables reflect the underlying truth of the situation they purport to represent is, of course, quite another topic. Whether the Buddha's early followers can really be likened to deluded children or not is debatable, but it is not a question I can enter into here. (373)

[2] At Reeves *Classic* 143, a father attempts to use coercion on his mentally unstable amnesiac son. The attempt fails: the son is utterly terrified and falls into a faint. The father then decides to resort to paternalist deception instead; he pretends to need stable workers and offers his son a job removing dung.

I propose to begin where Keown left off, and to consider a question closely related to the one he chooses not to answer. If we care about whether the ethical perspective of the *Lotus Sūtra* is substantively defensible, this is undoubtedly what we will have to do. But I shall not consider Keown's exact question, because I think it frames the issue in an unfortunate way. Those practicing the Way of the Disciples are not the only ones who may receive misleading teachings from the Buddha. Consider a beginning bodhisattva who forms the aspiration "I will become a Buddha," an aspiration without which one cannot really enter the Mahāyāna. But since there is no self, there is no "I" even now, much less a persisting self that will exist until the time of Buddhahood. Moreover, as Peggy Morgan points out, the *Lotus Sūtra*'s account of its own composition and historical setting, in a cosmic assembly of humans and nonhumans on Vulture Peak in the time of Śākyamuni, is literally, historically false and therefore, at best, itself an example of appropriate means (353-56). This account, of course, is addressed to any and all readers of the Sūtra. In order for the ethical perspective of the *Lotus Sūtra* to be defensible, it will have to turn out that all of us who are not fully awake are appropriate objects of paternalistic deception.

In order to explore whether that claim can be made good, it will help to examine the arguments that can be offered against paternalism. Now if we are looking for an ethical theory that is staunchly opposed to benevolent lying, we will find it in the writings of Immanuel Kant. Kant notoriously claimed that it would be wrong to lie to a murderer who demands information about the location of his intended victim; when he was taken to task for this claim by Benjamin Constant and others, he emphatically reasserted and defended it in a piece entitled "On a Supposed Right to Lie because of Philanthropic Concerns" (Ellington 63-67).

Kant has several main types of arguments available for rejecting the idea of benevolent lying; I will discuss one based on the first and

three based on the second formulations of the Categorical Imperative. The first type of argument is related to the Formula of Universal Law. Kant argues that the moral principle that forbids lying is universal, *a priori*, exceptionless, and holds in all cases whatsoever, no matter how disastrous the consequences of following it might be.[3] The other arguments are related to the Formula of Humanity. Kant can say that we have an obligation to respect the humanity of others, and that this obligation must be honored even at great cost. If we deceive others with the intention of benefiting them, we show disrespect for their rational nature and thus wrong them, even if the results turn out to be helpful.

Gene Reeves has helpfully pointed out how different the ethics of the *Lotus Sūtra* is from Kant's system of immutable, universal laws:

> [T]he Buddha provides four sets of prescriptions which bodhisattvas should follow ... But these should be understood, I think, not as commandments but more like counsel or rules of thumb. Principles, at least in the strongest sense, are eternal, God-given, or at least implanted permanently in the nature of things. The *hōben* of the *Lotus Sūtra*, in contrast, are provisional. Once used, they may no longer be useful, precisely because they were appropriate for some concrete situation. (*Appropriate* 387)

This kind of moral framework shouldn't really be described as relativist, since philosophers typically use the term "relativism" to refer to a theory in which whatever a particular individual or culture believes about ethics is automatically right for them. Instead, it would make more

[3] This is in fact the main tack he takes in his essay "On a Supposed Right ..." For example, he says: "But here one must understand the danger not as that of (accidentally) doing harm [*schaden*] but in general as the danger of doing wrong [*unrecht*]. And such wrongdoing would occur if I made the duty of truthfulness, which is wholly unconditional and which constitutes the supreme juridical condition in assertions, into a conditional duty subordinate to other considerations. And although by telling a certain lie I in fact do not wrong anyone, I nevertheless violate the principle of right in regard to all unavoidably necessary statements generally" (Ellington 67).

sense to think of the framework as consequentialist; what makes the Buddha's means appropriate is that they work, that they actually succeed in benefiting sentient beings.[4]

How credible is Kant's first line of argument? Many philosophers today would agree that Kant's main attempt to justify his own favored set of immutable principles, namely, the Formula of Universal Law, is a failure. The set of wrong actions, and the set of actions such that they wouldn't work if everyone did them, may overlap, but they pretty clearly don't coincide.[5] And Sidgwick's brilliant and searching examination of the "Morality of Common Sense" at the end of the nineteenth century should already have convinced us that we can't expect to extract a set of mutually consistent, exceptionless moral principles from the messy moral practices of our society (Book III).

So Kant's most promising strategies for rejecting the appropriateness of benevolent deception rely on a commitment never to treat humanity as a mere means, but always at the same time as an end. Now this kind of moral reasoning relies on a particular kind of philosophical anthropology, an account of human nature and of the self. For Kant, everything in the human personality that can be studied empirically is subject to, and determined by, causation. However, Kant asserts that there is, in addition to these determined phenomena, something else, a noumenal self, which is outside space and time. The noumenal self is fully and perfectly rational. When it chooses, it always chooses a morally permissible action. It is not possible for us to know, through the exercise of theoretical reason, that this noumenal self exists. But the noumenal

---

[4] See Reeves *Appropriate* 382. Reeves does not see the ethics of the *Lotus Sūtra* as exclusively consequentialist, since intentions matter too. But in Goodman 186-87, I show how to deal with this problem by moving to a subjective version of consequentialism. This kind of consequentialist theory can easily accommodate the role of intention in Buddhist ethics.

[5] As in the powerful example, "I will buy a toy train, but never sell one."

self can be an object of belief, or faith (Ger. *Glaube*.) Belief in the noumenal self is necessary, according to Kant, in order to make the moral life possible.

How should we regard the ascription of a rational noumenal self to all humans? After all, Kant does not want to deny that people sometimes act irrationally. We do so whenever we do something wrong. So what is he asserting when he ascribes a perfectly rational nature to all of us, and tells us that morality requires us to respect it? We could see this rational nature as a mere capacity, as a conclusive presumption based on our ignorance, or as an ideal of reason. Let's explore what each of these alternatives would involve.

One Kantian strategy would be to argue that the capacity for full rationality needs to be respected even in those people who rarely or never exercise it. So even in a case where we might know that the person in question will definitely do something irrational that will lead to misery, we still should not intervene deceptively or coercively to stop that person, out of respect for their capacity for rationality.

It strikes me that we will find this version of the argument from the Formula of Humanity plausible only if we have already been convinced that humanity, the capacity to set ends and think rationally about how to attain them, has unconditional moral value, whereas happiness does not. It is only by severely downgrading our understanding of the value of this person's happiness that we can see an intervention to protect it as wrong in virtue of infringing on a capacity that we know will not be exercised. Now Kant has a powerful argument for this understanding of moral value: the Regress to Humanity as an End. But although this argument has been persuasive to many, it can reasonably be rejected. In particular, Buddhists can and should reject Kant's argument. I make this case in my book, *Consequences of Compassion* (198-200), and I will not repeat it here.

But can we in fact have the kind of knowledge I have been assuming? Although Kant admits that people do make irrational choices, he also argues for severe limits to our ability to know the status of an action as rational or irrational. No one, not even the agent, can know for sure that an action was rational and free, since it may secretly have been done out of immoral motives. Moreover, although we might be able to know that someone else's action was morally wrong, we can't know what that other person takes her happiness to consist in, so we can't know whether an action that strikes us as merely imprudent was actually rational or not.[6] If it is impossible to know this information, then it makes sense that we would be obligated, in our relations with others, to conclusively presume that they are acting rationally, since we cannot know otherwise. And in this case, the generally accepted existence of human irrationality can offer no support to a paternalist action or policy. Instead, we are bound to respect the rationality and autonomy of others and allow them to make their own decisions. We may never try to manipulate their rationality with lies, especially for some alleged benefit to them, since this would be inconsistent with the dignity belonging to the free human nature which is always able to manifest in them. Paternalistic deception would mean that you are trying to control others for what you think is their good, instead of allowing them to express their own autonomy by freely choosing in the context of knowing the truth. But according to Kant, respecting their autonomy is precisely what morality requires, both in general and in this kind of case.

Finally, we could propose that we regard the ascription of rationality to others as an "ideal of reason." What would this involve? This way of working out Kant's argument is developed by Christine Korsgaard

[6] Kant's complex views about what we can and cannot know about the motivation and moral worth of actions are scattered across a number of places in his writings. See, for example, Infield 230.

in her essay "Two Arguments Against Lying."[7] Freedom and rationality are not properties whose presence or absence can be known through empirical evidence or *a priori* theoretical reasoning: as Korsgaard writes, "Actual conduct, then, does not provide evidence for or against freedom" (352). Instead, there is a practical and moral requirement to ascribe freedom and rationality to ourselves and others. Theoretical evidence can give us some guidance about which entities this requirement might apply to: we are unlikely to face a moral requirement to treat rocks as free and autonomous beings, for example. But in applying concepts of freedom to adolescents or to the mentally ill, there are inescapable cases of practical judgment that cannot be settled by any theoretical considerations: "We must *decide* who to count as a free rational being" (356). Moreover, according to Korsgaard, "The pressure of the moral law is towards treating every human being as a free rational being, regardless of actual facts" (352). Paternalist lies directed at normal adult humans would push directly against this moral pressure. On this interpretation of Kant's argument, in the absence of theoretical knowledge, we must make a choice; but certain choices, such as treating normal adults as appropriate objects of paternalist deception, are ruled out by the respect we owe to others.

What kind of account could Mahāyāna Buddhists, such as the people who composed the *Lotus Sūtra*, offer as an alternative to Kant's view? We could start, of course, with the doctrine of no self. Buddhists claim that we have a powerful, innate, largely subconscious commitment to the view that each of us exists as a real thing (Skt. *dravyasat*.) But this view is a mistake. Special, trans-empirical entities such as the soul or the noumenal self are utterly nonexistent fabrications, made up to enable philosophers to offer spurious justifications for this innate mistake. The individual person exists as a mere conceptual construction out of more

[7] The argument in Korsgaard is complex and subtle; I hope I do it justice here.

basic materials. These more basic materials may themselves, in turn, be the products of conceptual construction. On the view of the Abhidharma traditions, the process of reduction will eventually terminate in a collection of absolutely simple mental and physical phenomena. But according to the Madhyamaka view, we will never find anything that is not conceptually constructed. When we indicate, point out, think about, or refer to things, we can never avoid using categories that are created by our minds; and none of the entities we relate to or interact with can exist apart from these categories. Above all, our own existence has this same status: we do not exist as real, objective entities; we people exist only from a certain point of view.

The doctrine of no self has, alone, few implications relevant to our question. These implications emerge only when we conjoin it with a Buddhist analysis of the emotions. A number of Indian Buddhist texts present or presuppose a cognitivist view of emotions, closely analogous to that of the ancient Stoics. On this kind of view, emotions are not seen as non-representational, mute drives or urges; they are caused by, and saturated with, representational judgments. In the case of reactive emotions (Skt. *kleśa*) such as hatred, greed, desire, competitiveness, and pride, which dominate the minds of ordinary beings in cyclic existence, these judgments are comprehensively false. This view of the nature of emotions can be documented in several important Mahāyāna sources. In the *Holy Teaching of Vimalakīrti*, for example, the title character says: "Reverend Upāli, passions consist of conceptualizations" (Thurman 31). A number of passages that do not directly support cognitivism nevertheless indicate that reactive emotions are caused by false judgments. Vimalakīrti says that sickness "arises from the passions that result from unreal mental constructions" (Thurman 45). In Ārya Śūra's *Garland of Birth Stories* (*Jātaka-mālā*), similarly, we read: "Just as fire consumes the stick that kindles it, so anger destroys the man whose false notions give rise to it" (Khoroche 138).

Once they have arisen from mistaken conceptual judgments, reactive emotions then proceed to make the original problem worse by deluding and confusing us. This is why Śāntideva, in the *Introduction to the Bodhisattva's Way of Life*, says that anger is a "deceiver" (Crosby and Skilton 50). Ārya Śūra explains this view eloquently, again in relation to anger:

> He whose presence makes one blind, whose absence makes one clear-sighted—he stirred within me but did not escape me: Anger, I mean, who injures the man who harbors him … Anger makes him oblivious of the path to success lying open before him; instead, he stays away from it and so is deprived of fame and success … Usually he turns stupidly quarrelsome and too dull-witted to discern what is good for him and what bad. (Khoroche 137-8)

Although anger is particularly destructive in this regard, it's important to stress that the other reactive emotions are deceivers too. Desire, for example, often makes us overestimate just how good it would be for us to get the thing we want. Given that ordinary people spend most of their time oscillating from one reactive emotion to another, it follows that they spend most of their time being deceived by their own irrational patterns. This claim helps us understand the nature of the dream in which most of us live. Much of what we relate to, think about and care about consists of unreal projections created by our irrational thoughts and emotions. To see the world as it is, free from these projections, is a major component of what it is to be awake (Skt. *Buddha*.)

On Kant's view of ethics, no profound theoretical knowledge or spiritual insight is required to do the right thing. Absolutely everyone is able to know what is right and what is wrong, even in difficult cases, although most cannot articulate this knowledge as a universal principle.[8]

[8] Ellington 15-16: "Thus within the moral cognition of ordinary human reason we have arrived at its principle. To be sure, such reason does not think of this principle abstract-

By contrast, the *Lotus Sūtra*, along with other Mahāyāna Sūtras, describes humans as deeply confused, misunderstanding the world around them in ways that make it impossible for them to see the moral status and consequences of their actions clearly. Thus the *Sūtra of Innumerable Meanings*, considered part of the threefold *Lotus Sūtra*, tells us:

> All living beings, however, make delusory distinctions: weighing whether something is this or that; whether it is a gain or a loss. Bad thoughts come to them, producing a variety of evil actions. They transmigrate within the six states undergoing all kinds of suffering and harm, from which they cannot escape during innumerable billions of eons. (Reeves *Classic* 34)

From a Buddhist point of view, then, ordinary people are very closely analogous to the insane: "The world is a confusion of insane people striving to delude themselves" (Crosby and Skilton 94). Common sense tells us that a therapist dealing with a gravely mentally ill patient would be perfectly justified in going along with some aspects of the patient's delusional structure if doing so would make it possible for the patient to be more comfortable, less distressed, or less dangerous. This would be a clear case of weak paternalism, and we would see it as morally acceptable. But if the real truth is that we, the normal humans, are all crazy, then the same framework applies to us. Similar analogies could be drawn to the permissible treatment of the profoundly retarded, and, of course, of small children. This account allows us to make perfect sense of the way in which the *Lotus Sūtra* uses analogies that sound like weak pa-

---

ly in its universal form, but does always have it actually in view and does use it as the standard of judgment. It would here be easy to show how ordinary reason, with this compass in hand, is well able to distinguish, in every case that occurs, what is good or evil, in accordance with duty, or contrary to duty, if we do not in the least try to teach reason anything new but only make it attend, as Socrates did, to its own principle—and thereby do we show that neither science nor philosophy is needed in order to know what one must do to be honest and good, and even wise and virtuous. Indeed we might even have conjectured beforehand that cognizance of what every man is obligated to do, and hence also to know, would be available to every man, even the most ordinary."

ternalism to justify ethical prescriptions that would normally be seen as strong paternalism. On the Mahāyāna Buddhist view, the basic assumption of the distinction between the weak and strong versions of paternalism is a mistake: ordinary adult humans are not particularly rational. So what would look to most people like strong paternalism—namely, the use of deception on normal adults for their benefit—actually counts as weak paternalism, and as a result, will be easier to justify than we might have thought. As practiced by the Buddhas, paternalistic deception is an appropriate response to the profound delusions built into the mindset that creates the human realm.

Evaluating the Models

Between these two theoretical models—of humans as free, rational, autonomous, dignified moral agents who must be given the chance to make their own decisions with full information, and of ordinary people as confused, immature, irrational, struggling beings who can appropriately be deceived for their own good—which is more true to the way things are? Recent developments in psychology and related sciences have dealt severe blows to the Kantian framework and provided strong support to the views of the *Lotus Sūtra*. The social sciences have long been dominated by methodological rationalism, the project of understanding society on the assumption that people are rational in their decision-making. But today, methodological rationalism is in full retreat, routed by the experimental findings of behavioral economics and dismayed by the manifest wreckage of a financial system devastated by irrational choices.

One simple example of predictable human irrationality comes from an experiment designed by two professors of business at MIT, who set up an auction for Boston Celtics tickets. As Jonah Lehrer explains,

> Half the participants in the auction were informed that they had to pay with cash; the other half were told they had to pay with

> credit cards. [The experimenters] then averaged the bids for the two different groups. Lo and behold, the average credit card bid was twice as high as the average cash bid. When people used their Visas and MasterCards, their bids were much more reckless. (Lehrer 86)

This kind of experiment can go far towards explaining why, in the decade leading up to the present crisis, Americans have been borrowing more than they are saving and spending more than they can afford. Of course, there is no rationally justifiable explanation for the difference; the only possible explanation relies on the psychological unpleasantness the subjects experienced in actually parting with cash, as compared to how psychologically easy it is simply to charge a purchase.

Experiments by Dan Ariely have confirmed the existence of a striking phenomenon known as "arbitrary coherence." Ariely offered business students a list of five goods, including electronics, chocolates, and bottles of wine. He asked them to write down the last two digits of their social security numbers, expressed as dollar amounts, and then indicate whether they would be willing to pay that sum for each of the goods. He then held a sealed-bid auction, with real money, for each of these products. Astonishingly, the amounts that the students bid for the items were strongly correlated with the last two digits of their social security numbers. The effect was dramatic:

> The top 20 percent, for instance, bid an average of $56 for the cordless keyboard; the bottom 20 percent bid an average of $16. In the end, we could see that students with social security numbers ending in the upper 20 percent placed bids that were 216 to 346 percent higher than those of the students with social security numbers ending in the lowest 20 percent. (Ariely 28)

The students in this experiment were not totally lacking in economic logic. As Ariely writes,

> When we looked at the bids for the two pairs of related items (the two wines and the two computer components), their relative prices seemed incredibly logical. Everyone was willing to pay more for the keyboard than for the trackball—and also pay more for the 1996 Hermitage than for the 1998 Côtes du Rhône (Ariely 29).

However, the absolute amounts the students were willing to bid were profoundly influenced by an utterly irrelevant and non-rational factor: the random influence of their social security numbers. This experiment helps to illuminate both the enduring strengths and the long-overlooked limitations of economic theories based on methodological rationalism. It should also cast serious doubt on the idea that, when we are not acting wrongly, we make choices out of an intelligible faculty of practical reason.

Some empirical psychological evidence of irrationality helps to confirm the Buddhist view that reactive emotions, such as anger, cause us to have distorted and inaccurate beliefs. As one review article explains,

> [A]nger elicits a kind of "defensive optimism," in which angry people systematically de-emphasize the importance and potential impact on the self of the negative events (Hemenover & Zhang, 2004.) Finally, these effects appear even when angry subjects rate the likelihood of events for which anger is a predisposing factor. That is, even though chronically angry people are more likely to have cardiovascular problems (Fredrickson et al., 2000; Williams et al., 2000), experience divorce, and have difficulty at work (Caspi, Elder, & Bem, 1987), angry people rate themselves as significantly *less* likely than the average person to

> experience these problems (Lerner & Keltner, 2000, 2001). (Lerner and Tiedens 124)

Thus Śāntideva was right to think of anger as a deceiver.

One striking experiment offers further support for Buddhist claims about the irrational and counterproductive nature of anger and the desire for revenge. Carlsmith, Wilson, and Gilbert created an experiment in which undergraduate students who had just concluded a standard public goods-type prisoners' dilemma interaction were given the opportunity to spend some of their money to punish a participant who had urged them all to cooperate and who then herself defected. This participant was actually a computer, but the experimental subjects thought they were punishing another student who had cheated them. Some students were asked to predict how they would feel if given the opportunity to punish the free rider. Others went through the prisoners' dilemma trial but were not given the opportunity to engage in punishment.

Those subjects given the opportunity to predict their feelings stated that getting the chance to inflict punishment on the free rider would make them feel better. The students who were allowed to inflict the punishment reported that they would have felt worse had they not gotten the opportunity to punish. Both of these beliefs, however, were false. In fact, those students who were able to inflict punishment and did so were less happy than those who were not given the opportunity to punish. The experimenters concluded that this worsened affect was due at least in part to the fact that those who punished spent significantly more time thinking about the free rider's behavior (Carlsmith et al., 1320 and *passim*). This experiment gives us a clear example of a certain kind of emotion causing us to have a false belief. It also constitutes a plausible case of irrationality, since almost everyone has had the opportunity to punish others in large or small ways in the past, and yet the belief this

experiment exposed as false, namely that punishment leads to catharsis and thus to feeling better, persists in numerous human cultures.

These examples are in no way atypical. They could be multiplied at great length. Much of the work of experimental psychology over many decades has been devoted to finding forms of irrational behavior in humans. Each of them separately casts doubt on our self-image as rational choosers. Together, they render that self-image fundamentally untenable.

Just how does this experimental evidence bear on the two competing pictures of philosophical anthropology sketched earlier? It should be emphasized that the mere fact that people are sometimes irrational does not in any way contradict Kant's view. The problems for Kant arise from the claim that people are sometimes predictably irrational. If there are empirically discoverable, statistically robust patterns of irrationality that people in general tend to exhibit, then we can sometimes be in a position to know—not with certainty, but with the degree of reasonable confidence that is all we can ever expect in practical matters—that someone has been, is being, or will be irrational in making a certain decision. If this kind of knowledge is ever available to us, then the "conclusive presumption based on ignorance" argument cannot be successful.

What effect does this evidence have on the "ideal of reason" form of the Kantian view? Perhaps we find in Kant a conception of rationality so strong that no theoretical evidence could ever be sufficient to establish its presence. But to justify paternalist deception, the presence of this robust kind of rationality is not what we need to establish: we need its absence. And when an experimental subject is prone to arbitrary coherence, that subject is not being rational, either in a Kantian or any other sense. The experiments described above make it possible to know in advance that a certain group of people will not respond to a situation rationally; so theoretical reason can establish the premise that paternalist

deception needs to draw on. Of course, even if they will react irrationally and we know this, they might still retain the mere capacity for Kantian freedom; but if we have rejected the Regress to Humanity as an End, there is insufficient motivation for regarding respect for this capacity as so morally important as to override what we can know about these people's interests and the potential imminent threats to them. I conclude that none of the Kantian strategies I have considered are successful in defending an unqualified rejection of paternalistic deception.

The psychology experiments I have mentioned threaten not only Kant's critique of benevolent deception, but also his broader views about the self and about how human thinking works. They push us away from a picture of a perfect rationality arising from a mysterious noumenal realm, and toward a picture of humans managing to cope with the world around us through a wide variety of evolved heuristics, tricks, and kludges. This is the view of rationality developed by Dennett in his classic book *Elbow Room*. For Dennett, rational thinking arose through an unplanned, unguided process of natural selection, and it is implemented in an imperfect, buggy but immensely flexible and powerful physical machine, the brain. According to Dennett,

> The perfect Kantian will, which would be able to respond with perfect fidelity to all good reasons, is a physical impossibility; neither determinism nor indeterminism could accommodate it ... We are not infinitely but only extraordinarily sensitive and versatile considerers of reasons. (49)

We might question whether full awakening (Skt. *bodhi*) will be possible in this picture. It may well be that both karma and the possibility of awakening are a result of neuroplasticity, of the adaptability of the physical basis of our minds. But until we awaken, those who understand the limitations of our rationality may sometimes be in a position to manipulate

us deceptively for our own good; and if we understand the kind of beings we are, there are strong reasons to think that we should not object.

The Catch

Should we, then, happily ride off into the sunrise of a paternalistic utopia, in which our spiritual and political leaders regularly deceive and coerce us for our own benefit? If we are disturbed, or indeed frightened, by such a prospect, we should not try to rehabilitate Kant's philosophical anthropology; there really is nothing about us that should rule us out absolutely as objects of benevolent deception. Yet there is a severe problem with endorsing paternalism in practice. Paternalism is rightly exercised by adults over children; surely we cannot accept the idea that eight-year-olds could use paternalistic deception or coercion on other eight-year-olds. But if we are confused children, we may wonder whether the politicians and spiritual teachers we actually have are mature enough to function as our parents.

When we turn our attention to politics, it is blindingly clear that the answer is no. Just reflect for a moment on the American federal legislature, the Congress. While running for election to Congress, politicians attempt to tell the voters whatever they may want to hear. Ordinary citizens, in turn, typically cast their votes on the basis of racial or religious group membership, non-rational forms of identity, physical appearance, or inaccurate beliefs fostered by misleading television advertisements. Even if they make a sincere effort to choose in an informed and reasonable way, citizens in the voting booth are subject to the very same forms of irrationality and bias that plague them in their individual decisions. Moreover, the choice they usually face includes only two candidates with a realistic shot at winning; both of these candidates may be very seriously suboptimal, and the voters may have to select the lesser of two evils.

Once they arrive in Congress, legislators regularly, and with no sense of shame, make decisions in order to promote the interests of their own district, or even of their largest campaign contributors, at the expense of the public good. Actual laws, including those that will impose coercive sanctions on citizens, are written through a messy process of log-rolling and horse-trading. The politicians are focused on getting reelected, or possibly on the immensely lucrative jobs as lobbyists they may receive when they leave the Congress. Taking this system as a whole, does it seem like a source of benevolent, disinterested, authoritative guidance, free from any taint of irrationality? On the contrary, it seems like we may fervently wish to be left alone to make our own mistakes, rather than being subjected to coercive or deceptive control by such a comprehensively flawed and irrational system.

Perhaps the only form of paternalism it would make sense to tolerate from such a political system would be what Sunstein and Thaler have called "libertarian paternalism." Here the state intervenes to change the default option, thereby bringing about better social outcomes, but giving everyone the choice to do otherwise than what it suggests. So, for example, some people fail to sign up for IRA retirement savings plans that are available at their work, out of laziness, confusion, or weakness of will, thereby condemning themselves to a penurious old age. But if they were signed up automatically, as the default option, few would opt out. Government intervention to get their employer to change from an opt-in system to an opt-out system would be an example of libertarian paternalism. In this way, if someone has an immensely important reason to maximize their income now—say, an inordinately expensive but temporary medical crisis—that person could opt out of the system. Thus libertarian paternalism does not involve coercion or deception, and does not limit our freedom. Even the politicians we have, operating in the political system we have, can probably be trusted with this relatively weak but useful tool.

The *Lotus Sūtra*'s ethics of benevolent paternalism was probably never intended to apply to political leaders, except perhaps insofar as those leaders are themselves advanced bodhisattvas. But if the ethical views found therein are to have any relevance at all to practice, it must turn out that spiritual teachers can sometimes practice deception as a form of appropriate means. And perhaps it's plausible that they are better equipped to do so than the people who run our government. Certainly there is no shortage of people who claim to be accomplished spiritual masters but are actually charlatans and frauds. But those who are genuinely qualified to be Buddhist teachers must have some authentic experience of the way things really are, a form of experience that will make them more predisposed than most to use whatever capacities they have for the genuine, long-term benefit of those who accept their teachings.

It's important to emphasize again that the idea that the teacher can tell students what they need to hear at the time—which may not be what is actually true—is accepted, in some form, very broadly in the Buddhist world. Any tradition of interpretation that makes use of a distinction between teachings of provisional meaning (Skt. *neyārtha*, Tib. *drang don*) and teachings of definitive meaning (Skt. *nītārtha*, Tib. *nges don*) is thereby accepting that the Buddha does not tell everyone the whole truth, and may make statements that do not reflect ultimate truth but do help the listeners to move forward in their own spiritual journeys.

Now the *Lotus Sūtra* unambiguously regards this way of teaching as appropriate not just for the Buddha himself but for other spiritual teachers as well. Not only Buddhas, but also advanced bodhisattvas, are described as using appropriate means. (Reeves *Classic* 24) And this view about who can use paternalist deception is not restricted to the *Lotus Sūtra*, but is widespread in the Indian Mahāyāna tradition. We see this,

for example, in the often-repeated injunctions to teachers not to present the doctrine of emptiness to those who are not ready to understand it. So the *Lotus Sūtra* is simply presenting, in somewhat clearer, starker and more extreme forms, a view to which Mahāyāna Buddhists are independently firmly committed.

Indeed, contemporary spiritual teachers from Asian traditions have frequently made use of this form of teaching. A mild form of paternalist deception is illustrated in this charming story about the great Tibetan Buddhist lama, Kalu Rinpoche:

> A Canadian woman, one of his first Western disciples, visited him a few months before he passed away. He said, "Do you remember that first time we met? I asked your age, then told you that you had reached the ideal age and stage—neither too old, nor too young—to practice Buddhism."
>
> The woman had been forty-something at the time. Now, more than twenty years later, she replied, "Of course I remember! That meeting changed my life." She had let her career slide and had eventually entered retreat, first a three-year retreat, then a life devoted to contemplation and retreat that continues to this day.
>
> Rinpoche smiled a little impishly and said, "Well, right after you left, the next person who came to see me was a young woman in her early twenties, and I told her exactly the same thing!" (Zangpo 23)

A more serious, but still plausibly justified, example of this type occurred when Kalu Rinpoche was approached by a beginning student who asked to be taught the meditation practices of the Shangs pa bKa' rgyud lineage. He told the student, falsely, that the lineage had died out and no longer existed. If the student was not ready to receive these teachings,

but could benefit from other teachings Kalu Rinpoche could give him, then this seems like an acceptable example of appropriate means. That same student later wrote,

> Buddhists are very comfortable with the Buddha's and his followers' dedication to enlightenment above all else. For example, my Canadian friend and I profited immensely from Kalu Rinpoche's concern for us, which was far more important to him than an empty notion of honesty at all costs. (Zangpo 25-26)

Nevertheless, there is some reason to be wary of a view that allows teachers to deceive their students. To explain why, we might say that since real human teachers of today are not fully awake, they still have blind spots. And the trouble with your blind spots is precisely that you can't see them. If you give teachers permission to disregard otherwise binding moral rules—such as "don't lie"—for the greater benefit of others, you make it possible for them to promote the good much more effectively. But you also risk the possibility that their blind spots might turn out to coincide with the areas in which they're being given the power to break the rules. This scenario could lead to disaster; and as anyone familiar with the history of American Buddhism knows, this danger is not merely hypothetical.

So the permission to engage in paternalist deception is a dangerous gift, both practically and morally. But we should not conclude that there is safety in a complete prohibition of this type of deception. Suppose, for example, that emptiness is the way things are, but that those who hear the teaching of emptiness before they are properly spiritually prepared will have a tendency to misinterpret emptiness as nihilism and therefore to abandon all moral discipline, becoming both monstrous and miserable. These views seem to be central to much of Indian Mahāyāna. If they are correct, then a norm to the effect that a teacher must always

tell any student the whole truth about any subject would very likely be more dangerous than the *Lotus Sūtra*'s ethic of paternalist deception.

It's possible to agree that a qualified Buddhist teacher may tell different students contradictory things for their own benefit without accepting the stronger view that there are no rules at all that should bind such a teacher. One might hold, for example, that although teachers who are well on the way to awakening may lie to students, there is an absolute prohibition on teachers having sex with students. Alternatively, one might try to handle this very difficult and sensitive problem in a way consistent with the "no rules" view. Thus it would be possible to argue that although there is no exceptionless prohibition against sex between teachers and students, a request by a teacher to a student for sex is so unlikely to be a compassionate and helpful gesture, and so likely to be the expression of a selfish agenda, that the making of such a request should be seen by the student as *prima facie* evidence that this teacher can't be trusted with permission to break the rules. Paternalist deception, on the other hand, would not count as such evidence, although a student could certainly regard instances of hypocrisy or obviously selfish deception as impeaching the teacher's credibility.

I have argued that paternalist deception by teachers is a feature of the Mahāyāna quite generally, but some would argue that it is not a feature of Buddhism as such. In particular, Theravādins might argue that the *Lotus Sūtra*'s views on *upāya* are a contingent, historical product of the early Mahāyānists' unwillingness to admit to being innovators. We can hardly escape the conclusion that in the intellectual environment the early Mahāyāna practitioners faced, they did not have the option of claiming to be making progress—progress in ethical ideals, progress in meditation techniques, progress in philosophical formulations. They were forced to attribute deception to the historical Buddha precisely because they were forced to attribute their own, innovative teachings to

him. Theravādins, on the other hand, can and do claim that during his forty years of teaching, the historical Buddha never told a lie. So is their lineage exempt from the charge of paternalist deception?

This issue depends crucially on how we interpret the concept of conventional truth. It's well known that, according to both the Theravāda Abhidhamma and the Sanskrit Abhidharma traditions, statements about composite material objects and sentient beings are never ultimately true, but only conventionally true. As I have argued at length before, the most natural interpretation of this teaching, in the context of these particular traditions, is that ultimate truth is just truth, whereas conventional truth is something less—similar in some ways to fictional truth, and in other ways to approximate truth (Goodman "Vaibhāṣika"). If this interpretation is correct, then most of the Buddha's teachings were strictly and literally false. In the same way, a high school physics instructor, when lecturing on Newtonian physics, mostly says things that are strictly false—though they are highly useful, both pedagogically and practically, and approximately true.

Some scholars reject this account of conventional truth in the Abhidhamma traditions; so let's assume I'm wrong about this issue, and that for the Theravāda—as is the case for Madhyamaka—conventional truth actually is a kind of truth. Then a Theravādin could defensibly claim that the Buddha never used paternalist deception, in the strict sense, since he never told a lie. On the other hand, the historical Buddha, as he is presented in the Pāli Canon, was certainly willing to mislead students for their own good. He was therefore prepared to carry out a milder form of the same kind of teaching strategy.

To see this, let's consider a well-known text from the *Dīgha Nikāya*: namely, the *Tevijja Sutta*. This Theravādin scripture is famous for its epistemological critique of monotheism. The young Brahmin Vāseṭṭha asks the Buddha which of the Brahmin teachers of his day can

guide him to his religious goal, union with the god Brahmā. But the Buddha persuades him that none of these Brahmins is in any position to know how to be united with Brahmā, since they have not seen that deity and have no reliable evidence about him.

In this portion of the text, the Buddha's arguments would be reasonably familiar to someone like Richard Dawkins. But then, the dialogue takes what, to many readers, is a completely unexpected turn. The Buddha says:

> "Vāseṭṭha, it might be said that such a man on being asked the way might be confused or perplexed—but the Tathāgata, on being asked about the Brahmā world and the way to get there, would certainly not be confused or perplexed. For, Vāseṭṭha, I know Brahmā and the world of Brahmā, and the path of practice whereby the world of Brahmā may be gained."
>
> At this Vāseṭṭha said: "Reverend Gotama, I have heard them say: 'The ascetic Gotama teaches the way to union with Brahmā.' It would be good if the Reverend Gotama were to teach us the way to union with Brahmā, may the Reverend Gotama help the people of Brahmā!" (Walshe 193)

The Buddha then instructs Vāseṭṭha that to reach the world of Brahmā, a disciple must take ordination, keep moral discipline, develop meditative stability up to the first jhāna, and then thoroughly practice the Four Divine Abidings (Pāli *brahma-vihāra*), which are lovingkindness, compassion, joy, and equanimity.

This passage does not portray the historical Buddha as a liar: what he tells Vāseṭṭha is strictly and literally true. But Vāseṭṭha is being seriously misled. He thinks he is hearing instructions about the final religious goal, about freedom from cyclic existence—as he puts it at the beginning of the sutta, "the path of salvation that leads one who follows

it to union with Brahmā" (Walshe 187)—where Brahmā is undoubtedly to be understood as the supreme deity of the universe, in either a monotheist or a monist sense. But on the Buddha's own view, that's not what Vāseṭṭha is getting. The Buddha sees himself as giving Vāseṭṭha instructions on how to be reborn as a god in a long-lasting, but impermanent, heaven, a heaven that is squarely inside cyclic existence and in no sense represents a liberation from it. The Buddha could have told Vāseṭṭha this, but he didn't; perhaps he had good reasons to believe that, given the details of the young Brahmin's current psychological state, by practicing with the goal of union with Brahmā, Vāseṭṭha would make more progress than he would if he had been given a more accurate understanding of the nature of liberation in Buddhism. Or perhaps the Buddha knew that by working with the grain of Vāseṭṭha's worldview, rather than against it, he had a better chance of persuading the young man to engage in beneficial activities. On either understanding, the Buddha is promoting Vāseṭṭha's welfare by deliberately bringing it about that he will have an incorrect understanding of the nature of the goal of his own spiritual practice. This may not be paternalist deception in a strict sense, but it's certainly reasonably close.

This article has sketched a way of defending the paternalist views of the *Lotus Sūtra*, which as we have seen, points clearly and accurately to an uncomfortable, dangerous, but unavoidable normative feature of a Buddhist worldview. The requirement for teachers to adjust their teachings to the needs, interests and degrees of spiritual development of their various students, thereby opening themselves to the charge of dishonesty, is not an unfortunate innovation foisted on the Buddhist tradition by the *Lotus Sūtra*. It is an inevitable feature of a religious system which seeks to provide advanced students with guidance leading to a radical cognitive and affective transformation, while also functioning as a source of ethical values and moral discipline for large and populous so-

cieties. From a Buddhist perspective, it's hard to see how you avoid expecting the tradition to do both of those things.

I conclude that at least Mahāyāna Buddhists, and perhaps even Buddhists generally, are unavoidably committed to endorsing a form of weak paternalism; and given their understanding of human psychology, their view has implications that some would classify as strong paternalism. They are not required to endorse, and probably should not endorse, most forms of paternalist deception by political leaders, but only by certain advanced spiritual practitioners in their teaching activities. The ideal of a teacher who is utterly uninhibited and ruthless in using whatever means are available to bring students to awakening as quickly as possible can be both deeply appealing and deeply disturbing. But Buddhist practice is not about being comfortable. And there is no viable ethical objection to the conduct of such a teacher, always assuming that it actually works as advertised. In a world scarred and wounded by the results of greed, hatred and delusion, we need all the awakening we can get.[9]

Works Cited

Ariely, Dan. *Predictably Irrational: The Hidden Forces that Shape Our Decisions*. New York, NY: HarperCollins, 2008.

Carlsmith, Kevin, Gilbert, Daniel, and Wilson, Timothy. "The Paradoxical Consequences of Revenge." *Journal of Personality and Social Psychology* 95:6 (2008), pp. 1316-1324.

[9] Thanks to Gene Reeves, Luis Gómez, and all the participants in the 2010 International Seminar on the *Lotus Sūtra*—especially the late Bill LaFleur, who is sorely missed. Thanks also to my colleague Christopher Knapp for comments. Sigal Ben-Porath's presentation "Why Paternalism is Good For You," given at Binghamton University on Oct. 30 2008, showed me a new way to think about these issues, though I disagree with many of her substantive conclusions.

Crosby, Kate, and Skilton, Andrew, trans. Śāntideva. *The Bodhicaryāvatāra*. Oxford: Oxford University Press, 1995.

Dennett, Daniel. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press, 1990.

Ellington, James W., trans. Immanuel Kant. *Grounding for the Metaphysics of Morals*. Indianapolis: Hackett, 1993.

Goodman, Charles. *Consequences of Compassion: An Interpretation and Defense of Buddhist Ethics*. Oxford: Oxford University Press, 2009.

Goodman, Charles. "Vaibhāṣika Metaphoricalism." *Philosophy East and West* 55:3 (2005), pp. 377-393.

Infield, Louis, trans. *Immanuel Kant. Lectures on Ethics*. New York: Harper & Row, 1963.

Keown, Damien. "Paternalism in the *Lotus Sūtra*." In Reeves ed. 2003, 367-78.

Khoroche, Peter, trans. Ārya Śūra. *Once the Buddha was a Monkey*. Chicago: University of Chicago Press, 1989.

Korsgaard, Christine. "Two Arguments against Lying." In *Creating the Kingdom of Ends*, 335-62. Cambridge: Cambridge University Press, 1996.

Lehrer, Jonah. *How We Decide*. New York: Houghton Mifflin Harcourt, 2009.

Lerner, Jennifer S., and Tiedens, Larissa Z. "Portrait of the Angry Decision Maker: How Appraisal Tendencies Shape Anger's Influence on Cognition." *Journal of Behavioral Decision Making* 19 (2006,) 115-137.

Morgan, Peggy. "Ethics and the *Lotus Sūtra*." In Reeves ed. 2003, 351-66.

Reeves, Gene. "Appropriate Means as the Ethics of the *Lotus Sūtra*." In Reeves ed. 2003, 379-92.

Reeves, Gene, ed. *A Buddhist Kaleidoscope: Essays on the Lotus Sūtra*. Tokyo: Kosei Publishing Co., 2003.

Reeves, Gene, trans. *The Lotus Sūtra: A Contemporary Translation of a Buddhist Classic*. Boston: Wisdom Publications, 2008.

Sidgwick, Henry. *Methods of Ethics*. Indianapolis: Hackett, 1981.

Sunstein, Cass, and Thaler, Richard. 2003. "Libertarian Paternalism Is Not an Oxymoron." AEI-Brookings Joint Center for Regulatory Studies Working Paper No. 03-2. Available for download at the John M. Olin Program in Law and Economics Working Paper Series: http://www.law.uchicago.edu/Lawecon/index.html

Thurman, Robert, trans. *The Holy Teaching of Vimalakīrti: A Mahāyāna Scripture*. University Park: Pennsylvania State University Press, 1976.

Walshe, Maurice, trans. *The Long Discourses of the Buddha: A Translation of the* Dīgha Nikāya. Boston: Wisdom Publications, 1995.

Zangpo, Ngawang, trans. *Timeless Rapture: Inspired Verse of the Shangpa Masters*. Ithaca: Snow Lion, 2003.