

Argumentation (2008) 22:149–163  
DOI 10.1007/s10503-007-9069-4

---

## Reason, Irrationality and Akrasia (Weakness of the Will) in Buddhism: Reflections upon Śāntideva's Arguments with Himself

Tom J. F. Tillemans

Published online: 1 February 2008  
© Springer Science+Business Media B.V. 2008

**Abstract** Let it be granted that Buddhism has, e.g., in its logical literature, detailed canons and explicit rules of right reason that, amongst other things, ban inconsistency as irrational. This is the normative dimension of how people should think according to many major Buddhist authors. But do important Buddhist writers ever recognize any interesting or substantive role for inconsistency and forms of irrationality in their account of how people actually do think and act? The article takes as its point of departure a recurring theme in the writings of the 8th Century Indian Buddhist thinker, Śāntideva, who subjects his own behaviour and thought to minute scrutiny in argumentation with himself, only to be puzzled at his own seemingly irrational persistence in ways of thinking that he knows to be wrong and actions that he knows to be worse courses. The Buddhist's situation is profitably comparable to issues of akrasia, weakness of the will, that are taken up by Plato, Aristotle and many modern philosophers, including notably Donald Davidson and David Wiggins.

**Keywords** Śāntideva · Dharmakīrti · Akrasia · Inconsistency · Rationality · Ethics

What do we mean by rationality? We often tend to reach for a characterization in formal terms. Rationality can be seen as logical consistency, for instance. We can call someone irrational who

---

Versions of this article were presented to the Philosophy Department of National Taiwan University, the Buddhist Studies program in Smith College, the Sanskrit and Indian Studies Department in Harvard and the congress of the International Society for the Study of Argumentation at the University of Amsterdam. My thanks to many questioners and critics.

---

T. J. F. Tillemans (✉)  
Department of Oriental Languages and Cultures, University of Lausanne, Lausanne 1015,  
Switzerland  
e-mail: Tom.Tillemans@unil.ch

affirms both *p* and not-*p*. By extension, someone who acts flagrantly in violation of his own interests, or of his own avowed objectives, can be considered irrational (Taylor 1982, p. 97)<sup>1</sup>

What is special in incontinence is that the actor cannot understand himself: he recognizes, in his own intentional behaviour, something essentially surd (Davidson 1980, p. 42)

“Why do I do the things that make me sorry?”

(Lyle Lovett, “Who Loves you Better than I?”)

## 1 Irrationality and Akrasia

One of the ways to understand positions better on the nature and scope of reason is to look at the role, if any, they accord to irrationality in human thought and action. In the case of Buddhism, it is obvious that there is a great deal in the philosophical, doctrinal and ethical literature that testifies to highly developed norms of rationality. We regularly find scholastic authors of various Buddhist schools advocating the strict exclusion of contradiction as a canon of right reason: contradiction is said to be a *doṣa*, a fault in thought or speech, and a *nigrahasthāna*, a point of defeat in a debate. Let us speak of a kind of “standard conception of rationality”<sup>2</sup>, which is widespread in Buddhism, especially in its later Indo-Tibetan varieties heavily influenced by the Buddhist logical literature, but is also frequently presented in Western sources as a kind of rudimentary intuitive account of what it is to be rational. (It is what I have given above in the quotation from an article by Charles Taylor.) Buddhists sometimes formulate this standard conception in very strong terms: in a famous passage in the *Prasannapadā*, the sixth-century Indian author, Candrakīrti, says that when an opponent worthy to debate with ( $\approx$  rational) has been persuaded that his own position is riddled with contradictions, he *will* give it up. Someone who does not do so is said to be *unmattaka*, “out of his mind,”

<sup>1</sup> Note that “rational” is often used, *inter alia*, to describe what is simply based on reasoning, be that reasoning more or less sound or unsound, good or bad; or it is used to qualify beings that are endowed with the capacity to engage in reasoning. The adjective is also, of course, as in the passage from Charles Taylor, used in a normative and evaluative sense, to characterize thinking, speech, behaviour, etc. that are in keeping with *good* reasoning. In what follows, we are using the term in that latter sense. This is close to the use of the Sanskrit terms such as *yukta/nyāyā* and their Tibetan and Chinese translations *rigs pa*, *cheng li*, etc. Cf., however, van Eemeren and Grootendorst (2004, pp. 123–125), where the use of the term “rational” is stipulatively limited to the first sense and the term “reasonable” is reserved for the good uses of reason.

<sup>2</sup> There may be, e.g., in certain early Buddhist texts, and perhaps in certain Chan texts, non-standard conceptions which have a greater tolerance of inconsistency. That would need a separate treatment and I’m not going to delve into such issues here; accordingly my focus will be upon the “standard conception.” There is in any case no doubt that many Buddhists do profess a standard conception of reason where avoidance of contradiction is an explicitly invoked iron-clad rule: it is abundantly attested in their own philosophical theories, their canons of debate (*vāda*) and in their polemical refutations of adversaries.

“demented,” and hence not worth arguing with.<sup>3</sup> On this “standard conception”, then, at least a necessary condition for people to be rational is that they strive to be consistent; thus acquiescence in, or worse, willing espousal of inconsistency will accordingly be considered to be irrational.

Now, in Western thought there is an old conundrum that is often linked to the problem of irrationality: this is the philosophical problem, originating in Plato’s *Protagoras* and in Aristotle’s *Nicomachean Ethics*, of akrasia, i.e., weakness of the will, incontinence, the fact that people often seem to act contrary to what they think or know to be best. Potential examples of weakness of the will are familiar to all and abound, from giving free rein to anger that one knows to be senseless to going on and on in seamy relationships while having only deep pessimism about the probable outcomes. And there are interesting philosophical consequences if such weakness is acknowledged: many socio-economic or ethical theories, including some popular Buddhist teachings on ethics (see e.g. Ricard 2006), take it as more or less a given that people act in a way that they are convinced will further happiness and diminish pain, be it their own or that of others: if akrasia is to be taken into account seriously, pursuit of maximal utility will be at most an idealized or simplified model, one that may well be false in describing the actual complex features of people’s behaviour.<sup>4</sup>

Finally, akratic behaviour is often taken as having an epistemic counterpart too: we can know that that such and such a view is wrong or false, but nonetheless somehow believe in it just as if we thought it true (see e.g. Hookway 2001). Wishful thinking is an easy example; various sophisticated and seductive self-deceptions are more interesting potential cases of an epistemic version of akrasia. The problem is not just treated in works of philosophy or ethics: certain great works of literature, like Gustave Flaubert’s *Madame Bovary*, explore such forms of self-deception<sup>5</sup>; most of the famous case-studies of *mauvaise foi* which Jean-Paul Sartre so penetratingly analysed in *Being and Nothingness* are also arguably types of akrasia. In short, an akrates is convinced that *p* is best and yet acts as if he *in some sense* believes in *not-p*; he strongly thinks that *p* is true and yet also *somehow* accepts *not-p*. Akrasia is thus often seen as a type of irrationality, or dangerously close to irrationality, given the standard conception of a rational person not acquiescing in, knowingly accepting or promoting inconsistency.

Do Buddhists recognize any important role for akrasia in their account of how average, sane, people think and act? Let us from here on put the moral and epistemic together and just speak of the problem of akrasia taken in both fashions. Of course,

<sup>3</sup> Candrakīrti, *Prasannapadā*, ed. de la Vallée Poussin (1970: 15.9–10): *atha svābhyupagamavirodhacodanayāpi paro na nivartate, tadāpi nirlajjatayā hetuḍṛṣṭāntābhyām api naiva nivarteta / na commattakena sahāsmākaṃ vivāda iti*. “But if the opponent did not desist even when confronted with a contradiction in his own position, then too, as he would have no shame, he would not desist at all even because of a logical reason and example. Now, as it is said, for us there is no debate with someone who is out of his mind.”

<sup>4</sup> The falsity of theories that explain behaviour only in terms of maximizing utility/happiness is the consequence that David Wiggins draws from akrasia. See Wiggins (1998).

<sup>5</sup> Emma Bovary’s persuading herself that she would be swept off her feet by her lover and her elaborating numerous other romantic scenarios about love and even her own sainthood, all the while knowing they were not so, can be regarded as case studies of epistemic akrasia. Such is the analysis in Davidson (2004b).

Buddhists do recognize that people act and think in ways that they *later* realize to be wrong, or even hopelessly incoherent—indeed this is a sad fact of life that virtually everyone is forced to admit at some time or another. But do they recognize that people act and think akratically, i.e., believing things and acting in certain fashions, while *at the same time* in some sense aware of being inconsistent with their own better judgment and intentionally pursuing *worse* courses of action?

Issues of self-deception and epistemic akrasia have been taken up in connection with Chinese, Japanese and Indian thinkers by Roger Ames, Elliot Deutsch and others; Richard Hayes has also seen the interest of the problem in Buddhism, all the while acknowledging, as I do, that there is no treatment of it as an *explicitly* formulated philosophical theme in classical Buddhist writings.<sup>6</sup> My starting point and the work to which I return again and again is the *Bodhicaryāvatāra* of Śāntideva (7–8th century CE), where, I would argue, the problem underlies the discussion in several chapters, sometimes at almost every turn. The problem comes out especially urgently in Śāntideva’s self-analyses—instead of the non-conceptual quietism often associated nowadays with Buddhism, his practice clearly involves penetrating and relentless *argumentation with himself*.<sup>7</sup> And though the present analysis can be seen as an extended meditation on Śāntideva’s tortured auto-interrogation, there are other Buddhist contexts in which the problem seems to figure significantly. That said, a theme such as this is not primarily pursued by giving textual data from Sanskrit or other Asian language writings; it involves considerable interpretation and exegesis. The debt to the writings of Donald Davidson, David Wiggins and to the contemporary literature on akrasia and self-deception will become obvious.

## 2 Pure Rationality

To say, seemingly against the evidence, that it is utterly impossible that people act intentionally against their better judgment is to accept a version of what Donald Davidson has called a “doctrine of pure rationality”, whose first clear statement in the West goes back to Socrates in Plato’s *Protagoras* (see Davidson 2004b, pp. 174–175). Here is how Aristotle summarized that view in Book 7 of the *Nicomachean Ethics* (cited in Wiggins 1998):

One may be puzzled how a man with a correct view of a situation can be weak of will. For some deny that this is possible if he really knows what is the right thing to do. For if the knowledge is present, it is strange, as Socrates thought, for something else to overcome knowledge and manhandle it like a slave. Socrates was totally opposed to that view. He denied that there was any such thing as weakness of will. For knowing that it *is* best, nobody, he said, acts contrary to the best. If he does act contrary to the best, it must be through ignorance. This account of Socrates conflicts plainly with what seems to be the case and what people say...

<sup>6</sup> Although my analysis differs from that of Hayes (1996), I do agree with him on the importance, in this connection, of the Buddhist’s theory of personal identity and “modularization” of the self. See Sect. 7 below.

<sup>7</sup> On Buddhist quietism and the role for argumentation, including argumentation with oneself, see our introduction to this volume.

In short, a pure rationalist holds that when we pursue worse courses we don't actually know what is right, and if we did, we invariably would think and act correctly. If this was also the Buddhist position, then the Buddhist might well see people as being as exclusively rational as did Socrates. As we shall see in Sect. 4, there is a Buddhist account that goes in that Socratic direction, one that has had a certain success in popular presentations. Another textually supported account, taking seriously themes of some major Buddhist authors, does not.

### 3 Buddhist Sources Suggestive of Akrasia

David Wiggins began a long article on akrasia with the following astute observation: the existence of weakness of the will is not contested by ordinary non-theoretical people; it is a problem for theoreticians of rationality because they cannot see how it *could* exist and, especially, because they cannot see how it could be compatible with the other theories they espouse. Indeed I suspect that many non-philosophers, if asked, would reply (echoing Lyle Lovett) that we do sometimes, or even very often, intentionally do things that we know will make us terribly sorry, in short that we pursue worse courses and are deeply puzzled as to what reason we could have for doing so. Aristotle said that this was common knowledge, and it seems that Buddhist authors, when they are documenting or criticizing human foibles, take it as a clear fact too.<sup>8</sup> Although the Buddhist will differ with us on some of the details of what is or isn't painful or seamy, there is a recurrent argumentation in texts like Āryadeva's *Catuhśataka*, in Nāgārjuna's *Ratnāvalī* and in the *Bodhicaryāvatāra* of Śāntideva, to the effect that Buddhist practitioners, and indeed average people if they were to reflect a bit, actually know well that most of their actions are worse courses. Āryadeva (2nd–3rd century CE) for example, consecrates his first four chapters to a famous series of four “illusions” (*viparyāsa*) that are supposedly present in varying degrees in the troubled minds of Buddhist practitioners: thus Buddhists are convinced that things are impermanent but also tenaciously and wrongly hold them to be permanent, what is actually painful is held to be pleasant, unclean things are held to be clean, and selfless things are held to have selves (Lang 2003, 1986).<sup>9</sup> The logic in the treatment of these illusions is always the same: people are self-deceived; they know that permanence, cleanliness, selves, etc. are illusions and yet somehow think the exact opposite and act for the worse. Śāntideva adopts a first-person perspective and again and again engages in extensive arguments with himself, contrasting fundamental Buddhist truths and

<sup>8</sup> Cf. Wiggins (1998, pp. 239–240): “Almost anyone not under the influence of theory will say that, when a person is weak-willed, he intentionally chooses that which he knows or believes to be the worse course of action when he could choose the better course. ... But there are philosophers of mind and moral philosophers who have felt a strong theoretical compulsion to rewrite the description, rather than allow the phenomenon of weakness of will to appear as an incontrovertible refutation of the theories of mind or morality that they are committed to defend.”

<sup>9</sup> For the canonical schema of four *viparyāsa*, see *Abhidharmakośabhāṣya* ad V.8, translation de la Vallée Poussin, tome IV, p. 21.

norms which he (in some sense) knows with the opposite thought patterns of his own divided mind.

To take two typical verses:

Why, mind, do you protect this carcass, identifying with it? If it is really separate from you, then what loss is its decay to you?

O Fool! You do not identify with a wooden doll even when it is pure. So why do you guard this festering contraption made of filth? (Translation in Crosby and Skilton 1996, p. 39)

There are many other such descriptions of seemingly inexplicable odd behaviour and incoherence: Dharmapāla, for example, illustrates the “strangeness of the world”, with a list of ten “hard to fathom things” (Chinese: *nan ce*), some of which (such as people coveting money all the while renouncing the world) are suggestive of *akrasia* (Translated in Tillemans 1990, vol. 1, pp. 171–172). Other authors, especially the writers on monastic discipline (*vinaya*), describe in considerable detail the *bizzarries* of celibate individuals, who subscribe to the monastic code but nonetheless pursue (alas, often intentionally, I suppose) quite disturbing deviations from the rules. No need to elaborate further. In any case, there seems to be ample information in texts documenting people acting in a way that they knew or strongly believed to be wrong. In fact, there are more complex potential cases of *akrasia* where belief in errors is seen as inescapable and even necessary. These “positive” cases arise, not in ethical contexts, but rather in Buddhist metaphysics and philosophy of language. However, a discussion of these error-theories of universals and conventional truths, even though they may well be seen as a peculiar advocacy of epistemic *akrasia*, would take us too far into other matters (see Tillemans 2004).

Significant is that both Śāntideva and Dharmapāla express deep puzzlement about why people do the things they do against their better judgment. But what are these two writers puzzled about and what kind of an answer might they have expected to receive, if any? One quick, and unsatisfying, response would be to say that Śāntideva wasn’t actually puzzled at all—his questions were just repeated exhortations to himself to practice Buddhism. But this unfortunately amounts to eliminating the divided mind that Śāntideva reveals throughout so much of the *Bodhicaryāvatāra*: instead of a mind in conflict, we would have an essentially indolent figure who provides himself with a series of rhetorical goads. Charity demands that we take Śāntideva’s “Why?” more seriously.

Let me begin with what I think is the kind of answer that Śāntideva is probably *not* seeking. He and most other people who puzzle about why they choose worse courses (be they 8th Century Indian Buddhists or 20th Century country and western singers) are usually *not* seeking purely causal explanations.<sup>10</sup> Indeed, if the repeated “why?” (*kasmāt*) and “how is it possible?” questions that we find so often in the *Bodhicaryāvatāra* were to be taken as the search for a psychological explanation identifying causal factors that gave rise to odd behaviour and beliefs, a Buddhist should have little difficulty in coming up with answers. Those answers might be intricate, especially given Buddhism’s detailed discussions of karma, mental factors

<sup>10</sup> I am of course once again thinking of that great Texan akrates, Lyle Lovett.

(*caitta*), dependent arising (*pratītyasamutpāda*), and the other elaborations of mental causality found in Abhidharma or Pramāṇa literature, typically in connection with the question of the origin of suffering. But the answers would be forthcoming and relatively unproblematic.

In the West or in the East there is no shortage of causal explanation as to what makes people go against their better judgment: e.g., the passions constrain people to do what they don't want to do, much like external force; or the psychological force of better judgment is over-powered by the force of temptation; vigilance and mindfulness are too weak against the passions; or better judgment is no match for the strength of longstanding habits and ingrained tendencies to go against it, and so and on and so forth. Many of these causal explanations, and particularly the appeal to the causality of longstanding tendencies (*vāsanā*), are at one point or another invoked by Buddhists to explain psychologically how it happens that we do or think things we also know to be worse choices or errors. Indeed talk of deeply ingrained tendencies, or more exactly "beginningless tendencies" (*anādikālavāsanā*), is such a Buddhist cliché that if Śāntideva or another Buddhist had simply wanted a causal account of why people go against their better judgment, it would be hard to see why any puzzlement would have arisen at all. I think it is clear that a causal explanation along the lines of "Buddhist psychology" is not primarily what is being sought: it is just too readily available.

Nor would such Buddhist psychology settle the puzzlement: invoking that type of explanation largely misses the point. Generally, when people are perplexed about why they do the things they do, they are asking for reasons and not just mere causes, reasons that enable them to make sense of actions by seeing them as fitting into a pattern of thought and action that is rational and consistent. Such is, I maintain, also what seems to lie behind Śāntideva's "why?" The puzzlement comes from the following: while Śāntideva had what were for him good Buddhist reasons for not identifying with his body and not cherishing it, and he no doubt had all-too-human reasons for nonetheless protecting his body, he found himself with no reasons at all for pursuing a course of action (i.e., protecting his body) that he sincerely and deeply believed to be worse over one that he thought to be better. This is indeed the hallmark of akrasia: the actor cannot satisfactorily understand himself in reasoned terms; although he may be able to give causal accounts and may perhaps even be a determinist in maintaining that all his actions and thoughts are somehow caused by events in the past, he cannot give a reason that would entail that, all things considered, he should pursue what he knows to be the worse course.<sup>11</sup> This seems to correspond to Śāntideva's predicament, for the *Bodhicaryāvatāra* is above all an intense, almost obsessive, introspection of a thinker perplexed at having no reasons to justify his rejection of what he thinks best and true.

<sup>11</sup> The central idea is that of Davidson, encapsulated in one of the three quotations with which we began this article. See the elaboration in Davidson (1980, p. 42).

#### 4 Dismissals

So much for Buddhist observations and recognition of the oh-so-common phenomena of akrasia. *Theorizing* about how such phenomena are possible—i.e., making a place for the akratic in one’s theoretical account of human thought and action—is quite another matter, especially given the “standard conception” of rationality that Indo-Tibetan scholastic Buddhist thinkers promote. I think that many Buddhists, in reflecting theoretically, are reluctant to acknowledge that people’s thought and actions can be as akratic as they seem to be and feel that appearance of akrasia needs to be explained away. I can see three basic theoretical strategies that are in accord with broad lines of Buddhist scholastic texts and popular teachings, two that would dismiss all akrasia as only apparent and one that would acknowledge it and even provide the beginning of an account of how it is possible. Each involves, *inter alia*, an interpretation of the key Buddhist idea that ignorance (*avidyā*) underlies moral and epistemic faults. We begin with the two dismissals.

One potential Buddhist strategy to dispel appearance of akrasia can be elaborated as follows: in the Śāntideva-style examples, four illusions and other such cases, *a Buddhist is not actually in conflict with his own clear and better understanding, because the latter is for all intents and purposes nonexistent, or too fleeting, too vague, and hence cannot enter into any such conflict*. In short, there is no inconsistency between the principles that a Buddhist holds and the opposing beliefs that seem to guide how he actually behaves, because there is no real understanding of the relevant Buddhist principles at all. The would-be akrates might possibly pay lip-service to those ideas and even claim that he masters them, but in fact he doesn’t understand them in a way that could be said to be in genuine conflict with anything. The culprit, to put it in Buddhist terms, is the mental factor of ignorance (*avidyā*) that is at the root of the errors and misdeeds that keep sentient beings bound to the cycle of suffering and reincarnation, i.e., *saṃsāra*; ignorance clouds the mind so that understanding is weakened. Applied in a thorough-going fashion this dismissal leads to a doctrine of pure rationality à la Socrates: no akrasia, no inconsistency, just the all-present “darkness of ignorance” (*avidyāndhakāra*): we’re left with people (Buddhists or otherwise) who are rational and consistent, but who, in their dullness, don’t actually grasp the points at issue—a not very flattering depiction of serious, profound thinkers, like Śāntideva and others, who have consecrated their lives to understanding Buddhist philosophy.

Of course, this image of *avidyā* as *andhakāra*, gloom-making, is amply attested in Buddhist canonical literature, as are numerous other similar metaphors: ignorance as a cloud, an obscuration, a veil, fog, an eye disease. Preponderance of such imagery might seem to suggest that the ignorance that leads us astray was *invariably* interpreted as an unknowing that deprived people of better judgment, making them into “children” (*bāla*) and “fools” (*mūrkhajana*). And if that was all there was to it, then the “ignorance dismissal” would indeed put a fast end to any debate about akrasia, self-deception, and irrationality. To use an Aristotelian image, people invoking ethical or philosophical principles would be like drunkards reciting verses of Empedocles that they didn’t understand: such people might be stupefied, but they would not be akratic, nor would they be irrational in any interesting way.



The second potential dismissal of akrasia is a variant upon the first. It could be summarized as follows: *people don't act against what they think best; due to ignorance and delusions like self-grasping, they not only don't understand what is best, but erroneously believe that actions based on stinginess, indulgence in desires, and so forth are not the worse courses, but are the best courses for them in that they will maximize utility. Similarly for cases suggestive of epistemic akrasia and self-deception: people do not actually believe that permanence, the existence of selves, etc., are illusions, but believe they are real facts.* In short, people may again pay lip-service to Buddhist moral and philosophical principles, but they believe in *something else* that will maximize their own utility and that is why they pursue it. The first type of dismissal was based on people simply not understanding Buddhist principles and thus not adequately understanding what the best course of action was. The second had the additional feature that people would not actually believe that the worse course was worse, because ignorance made them take the false and worse for the true and best. The consequence of the second is in any case another variant on the Socratic principle of pure rationality: people would never intentionally pursue what they genuinely knew to be worse courses and would never believe what they knew to be false.

Who subscribes to pure rationality and its resultant dismissals of semblances of akrasia? It is above all to be found in contemporary interpretations of Buddhism as an “art of happiness,” or even a “science of happiness,” interpretations that often advance a position that goes strikingly in the direction of pure rationality and hence a dismissal of phenomena like akrasia. Indeed, what seems presupposed as a philosophy of action in this burgeoning literature, and is admittedly articulated with varying degrees of clarity and explicitness, is something that we can formulate along the following lines: sentient beings, by nature, will act in a way that they think maximizes their happiness; their pursuit of happiness, however, is misguided and regularly leads them to unfortunate actions because they are so deeply ignorant about what happiness really is and what leads to it.<sup>12</sup> This approach has no room for akrasia and precludes that there are people whose minds are divided between genuine better judgments of what is valuable and intentional pursuit of courses known to be worse.

One thing that needs to be made clear: Buddhism as a rational “art/science of happiness” is itself a thoroughly modern interpretation of Buddhism, an attempt at elaborating a Buddhist ethical philosophy that applies a broadly speaking utilitarian approach as a master-argument to decide courses of action. Why such a philosophy has been promoted and why it might seem attractive is beyond the scope of this paper, but would involve a number of considerations about how certain Buddhists have seen their religion's possible role in the world and its compatibility with modernity. In any case, the textual sources in Buddhism significantly underdetermine

<sup>12</sup> See Ricard (2006, pp. 26–27): “However we go about looking for it, and whether we call it joy or duty, passion or contentment, isn't happiness the goal of all goals? ... Anyone who says otherwise doesn't really know what he wants; he is simply seeking happiness under another name.” An appeal to the self-evidence of the fact that people do what they do because they rationally (but misguidedly) calculate it will maximize their happiness is clearly found in a recent article by the American monk, Thanissaro Bhikkhu (2006). On p. 43 we are provided with the following diagnosis of what underlies attachment to things like alcohol: “... in your calculation, the immediate pleasure derived from the alcohol outweighs the long-term damage it's doing to your life. ... We're attached to things and actions, not because of what we think they are but because of what we think they can do for our happiness.”

this currently promoted picture of Buddhist ethics as utilitarian.<sup>13</sup> Nor is there, as far as I can see, clear evidence in classical Buddhist literature in favour of the background principle that people will invariably act in the way they think, all things considered, to be most conducive to their happiness/utility. On the contrary, as we've seen in Sect. 3, we seem to have ample evidence that influential Buddhist authors recognize that certain people in pursuing worse-courses do not act in that way: they act in a way that they themselves understand very well leads to less overall happiness for themselves, and *a fortiori* less happiness for others. Indeed, following the *Bodhicaryāvatāra* that is virtually the rule for thoughtful people embarking upon Buddhist practice. In what follows, we will seek to develop interpretations of Buddhist theoretical notions, like *avidyā* and others, which make such conflict possible and do not *explain away* those important phenomena in favour of pure rationality.<sup>14</sup>

There are, in Buddhist philosophical literature, the makings of a more complex and rich theoretical account of akrasia, one that: (a) interprets ignorance as an active defiled intelligence, rather than a mere unknowing; (b) allows for conflicting networks of reasoning in different sub-systems, or different “compartments”, of consciousness; (c) develops a theory of personal identity and responsibility so that compartmentalized mental states are still attributable to one agent. All these themes are present in writings of major Mahāyānist Buddhist authors and in combination they give an account of human thought and behaviour in which there will be a considerable role for the all too human phenomena of akrasia in its various forms. In the remainder of this article, I'll sketch out some observations on each such theme in Buddhism. Hopefully the outlines of a more thorough treatment will become visible.

## 5 Ignorance as Defiled Intelligence

If a Buddhist wishes to defend pure rationality come what may, no doubt the easiest route for him, as we saw, is to generalize the role of ignorance as a simple unknowing: there are never any real cases of akrasia; the would-be akrates always turns out to be ignorant of what is best and true. While many Buddhists did take ignorance in that way, that is, as a non-understanding (*apratipatti*), it is important to note that many other major Buddhist philosophers, like the 7th Century author, Dharmakīrti, and the author(s) of the *Yogācārabhūmi*, did *not* see the ignorance that underlies *samsāra* as a mere absence of understanding, but as a very tenacious and even reasoned misunderstanding (*vipratipatti*). They explained it as being the wrong view that reifies persons, i.e., *satkāyadr̥ṣṭi*, or a “defiled intelligence” (*kliṣṭā prajñā*). I quote a relevant passage from Dharmakīrti's *Pramāṇavārttika* and *Svavṛtti* (auto-commentary) in full:

What then is the source of these [moral] faults so that they are abandoned due to the repeated cultivation of its antidote?

<sup>13</sup> On Buddhist ethics as virtue ethics rather than utilitarianism, see Keown (2001). Buddhists' narratives of dilemmas would seem to show that they, like most of us, are faced with what Thomas Nagel calls a “fragmentation of value”, i.e. competing values that come into conflict, and no single criterion to decide everything (Nagel 1979, Ch. 9).

<sup>14</sup> Cf. Wiggins' remarks cited in footnote 8 above. To persist in saying that Śāntideva just cannot actually know or believe such truths is tantamount to a refusal to allow challenges to a faulty background theory.

The genesis of all the [different] kinds of faults is due to the false view that reifies the person (*satkāyadarśana*). This constitutes ignorance (*avidyā*); the attachment to the [I and the mine] is based on that. From that [attachment] come hatred and [all] the rest [of the moral faults]. 222 Indeed if someone sees that there is no I and no mine, then, lacking [such] a belief, he will not be attached to anything. Nor will a detached person feel hatred for anything, for no-one has any [hatred] for what is harmless to himself and his possessions or prevents harm [to them]. Thus, the view reifying the self, which came about due to repeated cultivation of [a view] of the same general kind, leads to the view of there being a “mine”, and these two [views] to attachment to the [I and mine], and [finally] that [leads] to hatred and [all] the rest [of the moral faults]. That is why all faults arise from the false view that reifies the person. And it is precisely that [view] that we term ‘ignorance’ (*ajñāna*).<sup>15</sup>

In short, the necessary condition for the perpetuation of the world of suffering, i.e., *samsāra*, is ignorance taken as being a defiled intelligence reifying persons: this leads to the mistaken belief/apprehension that there is a substantial self and its possessions, to the related attitudes of self-protection and aggression, and finally, when these attitudes prevail, to the resultant worse-course behaviour, like stinginess, anger, and indulgence in desires, etc. Other thinkers, of a Madhyamaka orientation, like Candrakīrti and Tsong kha pa, also speak of active, intelligent, misconstruing, rather than just dullness: thus *satyābhimāna* (the attachment that things are real), grasping at the real (*bden ‘dzin*), etc. are underlying necessary conditions for our choosing worse alternatives.<sup>16</sup>

For philosophers who see *samsāra* as thus rooted in an active and defiled type of intelligence, rather than a simple *unknowing* of what is best, the facile Socratic “ignorance defense” for pure rationality is not readily available: there is, rather, a clash between strongly held ways of thinking. This leaves the route open for Buddhists to say that well-cultivated practitioners, like Śāntideva, could *know* they are wrong in pursuing stinginess, anger, and other passions, and in believing in permanence, cleanliness, selves, etc. They could know those things, though their defiled intelligence strongly believes the opposite and leads them to act upon what they know to be wrong.

## 6 Compartmentalization and Consistency

When we speak of an akrates going against his own better judgment, we easily attribute to him a type of inconsistency. But does that mean that an akrates who knows that generosity is the best course of action but nonetheless pursues his own stingy self-interest explicitly thinks to himself that one and the same action is both good and not good? Does he endorse a conjunction of a proposition and its negation? This seems quite unlikely and would be perilously close to what Candrakīrti calls simply “being

<sup>15</sup> *Pramāṇavārttika* I, 222 and *Svavṛtti* (ed. Gnoli 1960, p. 111): *kaḥ punar eṣāṃ doṣānāṃ prabhavo yatpratīpakṣābhyāsāt prahīyante/sarvāsāṃ doṣajātīnāṃ jātiḥ satkāyadarśanāḥ/sā avidyā tatra tatsnehas tasmād dveṣādīsambhavaḥ/222//na hi nāhaṃ na mameti paśyataḥ parigrahaṃ antareṇa kvacit snehaḥ/na cānanurāgiṇaḥ kvacid dveṣaḥ/ātmātmyānuparodhīny uparodhapratighātīni ca tadabhāvāt/tasmāt samānajātīyābhyāsajam ātmadarśanam ātmīyagrahaṃ prasūte/tau ca tatsnehaṃ sa ca dveṣādīn iti satkāyadarśanajāḥ sarvadoṣāḥ/tad eva ca ajñānam ity ucyate/.*

<sup>16</sup> For references to the two senses of ignorance in Candrakīrti’s *Pañcaskandhaprakaraṇa* and to “defiled intelligence” in the *Yogācārabhūmi*, see Seyfort Rugg (2002, n. 148). For *satyābhimāna*, see Candrakīrti’s *Madhyamakāvataṛabhāṣya* ad *Madhyamakāvataṛa* VI.28, translation de la Vallée Poussin (1910, pp. 304–305).

demented”, “out of one’s mind” (*unmattaka*), rather than being a sane worldling in a complex tension between better judgment and defiled forms of intelligence.

Now for the Buddhist, as for Donald Davidson and many other theorists about akrasia, the way back from such a radical form of irrationality is compartmentalization of mind: instead of one and the same mind or subject holding two opposing ideas or networks of ideas—a puzzling phenomenon indeed—there are supposedly many semi-autonomous cognitive sub-systems each having their own beliefs, with no substantial person linking them together.<sup>17</sup> I think the most plausible reconstruction of the position found in *Pramāṇavārttika* and other key texts<sup>18</sup> would be as follows: ignorance-qua-defiled intelligence is regularly depicted in Buddhist literature as creating a network of reasonings (e.g., “This is me”, “This is mine”, “This is hence essential to my happiness”, etc.) with a great deal of internal cohesion and leading to a coherent series of decisions to act in certain fashions (e.g., “I should defend this”, “I will be aggressive to someone who threatens me in this respect,” etc.). It is this network of reasonings and actions that leads to conflict with a competing ensemble of beliefs that there is no real I, no need to defend it, etc. The akrates, Buddhist or otherwise, could be depicted as caught between networks of ideas leading to him endorse *p* from the perspective of one such network and *not-p* from the other, but there would be no perspective from which he endorsed both *p* and *not-p*. The inconsistency that he endorses, if any, is thus of a weaker “non-adjunctive” sort; and that need not be the same as believing the conjunction *p and not-p*; he need not endorse the considerably stronger inconsistency that something is both good and not good.<sup>19</sup>

## 7 The Self

Of course, this is not the end of the story: there are other major philosophical hurdles that the Buddhist must cross. Let me sketch out two of them.

First, seeing the mind as a number of sub-systems is, as is often stressed in the East and in the West, particularly vulnerable to the critique that thoughts, desires, intentions and so forth would not end up being attributed to the whole person, i.e., to a unified subject that we refer to by “I”.<sup>20</sup> Buddhists were certainly aware of that

<sup>17</sup> These components are mental elements (*dharma*) classified into “minds” (*citta*) and “mental factors” (*caitta*)—this is the sort of thing that is explained in extraordinary detail in scholastic texts like the *Abhidharmakośa* and so many others of the Abhidharma genre. Fragmentation of the mind into such components is also explained, or presupposed, in *Yogācāra* and *Madhyamaka* Buddhism, even if the positions on the ontological status of such elements will differ from that of the Abhidharma. Compartmentalization is, in short, part of basic Buddhism.

<sup>18</sup> See the passage from *Pramāṇavārttikasvavṛtti* quoted in Sect. 5.

<sup>19</sup> Cf. Davidson (2004b, p. 217): “The distinction we need here is between believing contradictory propositions and believing a contradiction, between believing that *p* and believing that *not-p* on the one hand, and believing that (*p and not-p*) on the other.” I have taken up Buddhist uses of non-adjunctive inconsistencies in some detail in “How do Mādhyamikas Think? Notes on Jay Garfield, Graham Priest and Paraconsistency,” forthcoming in the papers of the BILAP (Buddhism in Logic and Analytic Philosophy) conference in Cambridge, 2005, ed. Jay Garfield, Mario D’Amato and Tom Tillemans. One could even weaken the inconsistency further by saying, with Wiggins, that perspectives may embody incommensurable values.

<sup>20</sup> The objection is that of the Brahmanical schools. But see also Davidson (2004a, p. 171) for a similar objection.

problem and had their makings of a reply. The main thrust of this philosophy, either in the Abhidharma, in Yogācāra or in Madhyamaka, is always that talk of the unity of the knowing subject is only a customary truth (*saṃvṛtisatya*), one due to worldly transactions (*vyavahāra*)—ascription of mental states to “whole persons” thus has to be explained without ascribing these states to a real I that owns them. This is not the place to go into an extensive treatment of Buddhist views on personal identity—others, like Mark Siderits, have taken this up in detail, looking especially at the Buddhist-inspired reductionism of Derek Parfit as well as Madhyamaka views on selflessness (Siderits 2003). Suffice it to say here that *some* such account of the unreality of substantial selves would seem to be indispensable if the Buddhist compartmentalist, or the thoroughgoing Davidsonian for that matter, is to allow for phenomena like akrasia.

Second, even if one does follow the Buddhist in relegating the unifying “I” to mere customary truth, what is also crucially necessary for philosophers who accept compartmentalization is an account of some kind of “panoptical” perspective, so that some or many of the mental sub-systems are transparent and accessible to each other.<sup>21</sup> Lacking some such transparency and integrating perspective, a thinker like Śāntideva would end up *too* divided to be perturbed at all by the conflicts between his mental states. Now, it may be that certain Buddhist philosophers, like Dharmakīrti, can fulfill that requirement for integration with their idea that every mental state is also aware of itself, and that this “reflexive awareness” (*svasaṃvedana*) can provide the needed panoptical perspective. However, it is not sufficiently clear to me how his account would work here, for the reflexive awarenesses of mental states are also taken to be individual and thus quite separate from each other. Curiously enough, it was not actually the Buddhist followers of Dharmakīrti who explicitly used “reflexive awareness” to integrate otherwise overly separate mental components. It was their philosophical cousins and rivals, the Kashmiri Shaivites, Utpaladeva and Abhinavagupta, who accepted much of the basic scheme of Dharmakīrti’s partitioning of the mind, but then invoked a unitary “reflexive awareness” to give the panoptical perspective that would permit separate components to communicate (see Torella 2007).<sup>22</sup> The idea may hold promise for rendering the divided mental sub-systems of the akrates transparent too.

## 8 Final Remarks

Buddhist philosophy of mind, in addition to compartmentalization, frequently advocates a type of stratification of the mental into conscious and unconscious strata. Indeed, a number of years ago the late Gadgin Nagao published an article in

<sup>21</sup> The ideas of a “panoptical scanner” and “transparency” are developed by Rorty (1988).

<sup>22</sup> Utpaladeva takes as an example the case of the two separate cognitions in Dharmakīrti’s account of non-perception (*anupalabdhi*) of a jar, i.e., the cognition of an empty place, like a table, and the cognition of it being devoid of jars. Lacking a type of panoptical scanner, i.e., a unitary *svasaṃvedana* that grasps both, the first cognition could not lead to the second. The example is generalized by the Shaivite to show that all separate cognitions can “communicate” with each other and enter into networks because of this illuminating feature of consciousness.

the *Felicitation Volume for Jacques May* in which he tried to look at the Buddhist *ālayavijñāna* (the storehouse consciousness) as a kind of unconscious in depth psychology, underlying the conscious mind, or *pravṛttivijñāna* (Nagao 1992). Recently William Waldron has taken a similar tack with a book on the *ālayavijñāna* (“fundamental consciousness”, or more literally “storehouse consciousness”) entitled *The Buddhist Unconscious* (Waldron 2003). Whether we see Buddhist enlightenment in Yogācāra terms as a “revolutionary perspective” (*parāvṛtti*) on the *ālayavijñāna* or along the Madhyamaka lines of an elimination of sentient beings’ “inborn grasping at things being real”, progress on the Buddhist path invariably involves bringing error to the ever-increasing light of day. What place does akrasia have here?

To put things in rough and ready terms, some thoughts—like the deepest, or innate (*sahaja*), grasping at a real self and the subtle forms of grasping at permanence and other illusions (*viparyāsa*)—could indeed be called “unconscious” in the sense that though a person thinks that *p*, he might not have the second order understanding that he thinks *p* (McGinn 1979), or if he did, he might not have any understanding of the extent of its hold on his conscious life. When these thoughts are unconscious, akrasia would *not* occur in connection with them, for there is no question at this time of knowing what is best but nonetheless consciously/intentionally choosing to do or think otherwise. However, as a practitioner becomes more and more familiar with his own mind and understands better its illusions, it seems that akrasia-like conflict becomes acute. Indeed, at the first stages (*bhūmi*) of a Mahāyānist’s spiritual advancement, the post-meditative (*rjes thob; pṛṣṭhalabdha*) consciousness is said to be under the influence of tendencies to “grasp at the real” (*bden ‘dzin*) and to give rise to passions, i.e., what the scholastic refers to as *klesāvaraṇa*, or “obscurations due to the passions”. Even after their elimination after the seventh *bhūmi*, the practitioner is still subject to a subtle obscurations, viz., that “phenomena appear [to him] as real” (*bden snang*). These are errors earlier realized to be wrong during meditative absorptions, but that continue to reappear when the practitioner interacts with the world. The tension between meditative perspectives and post-meditative reentry into the world thus remains, albeit in ever more attenuated forms, until full buddhahood. What is disturbing—just as I think it was for Śāntideva—is that some type of akrasia in all its puzzling urgency seems to be the price of spiritual progress along the Buddhist path.

## References

- Candrakīrti. 1907, 1910, 1911. *Madhyamakāvatāra* and *Madhyamakāvatārabhāṣya*, translated by Louis de la Vallée Poussin, In *Madhyamakāvatāra*, Introduction au Traité du Milieu de l’Ācārya Candrakīrti, avec le commentaire de l’auteur, traduit d’après la version tibétaine. Louvain: Le Muséon.
- Candrakīrti. 1970. In *Prasannapadā*, ed. L. de la Vallée Poussin. Osnabruck: Bibliotheca Buddhica.
- Crosby, K., and Andrew Skilton. 1996. *Śāntideva. The Bodhicaryāvatāra. A guide to the Buddhist path to awakening*. Oxford: World’s Classics, Oxford University Press.
- Davidson, D. 1980. How is weakness of the will possible? In *Essays on actions and events*, ed. D. Davidson. Oxford: Oxford University Press.

- Davidson, Donald. 2004a. Paradoxes of irrationality. In *Problems of rationality*, ed. D. Davidson. Oxford: Oxford University Press.
- Davidson, Donald. 2004b. Who is fooled. In *Problems of rationality*, ed. D. Davidson. Oxford: Oxford University Press.
- Dharmakīrti. 1960. Pramāṇavārttika and Pramāṇavārttikasvavṛtti. In *The Pramāṇavārttikam of Dharmakīrti*, ed. R. Gnoli. Rome: Serie Orientale Roma
- Hayes, Richard P. 1996. Ritual, self-deception, and make-believe. A classical Buddhist perspective. In *Self and deception. A cross-cultural philosophical enquiry*, eds. Roger T. Ames and Wimal Dissanayake. Albany, NY: State University of New York Press.
- Hookway, C. 2001. Epistemic akrasia and epistemic virtue. In *Virtue epistemology, essays on epistemic virtue and responsibility*, eds. A. Fairweather and L. Zagzebski. Oxford: Oxford University Press.
- Keown, D. 2001. *The nature of Buddhist ethics*. New York: Palgrave, St. Martin's Press.
- Lang, Karen C. 1986. *Āryadeva's Catuḥśataka*. Indiske Studier VII. Copenhagen: Akademisk Forlag.
- Lang, Karen C. 2003. *Four illusions, Candrakīrti's advice to travelers on the bodhisattva path*. Oxford: Oxford University Press.
- McGinn, Colin. 1979. Action and its explanation. In *Philosophical problems in psychology*, ed. N. Bolton. London: Methuen.
- Nagao, Gadjin M. 1992. The Yogācāra cognition theory and depth psychology. In *Etudes bouddhiques offertes à Jacques May*, eds. K. Mimaki, J. Bronkhorst and T. Tillemans, 1. Asiatische Studien/Etudes asiatiques 46.
- Nagel, T. 1979. *Mortal questions*. Cambridge: Cambridge University Press.
- Ricard, Matthieu. 2006. *Happiness. A guide to developing life's most important skill*. New York: Little, Brown.
- Rorty, Amélie Oksenberg. 1988. The deceptive self: Liars, layers and lairs. In *Perspectives on self-deception*, eds. B.P. McLaughlin and A.O. Rorty. Berkeley: University of California Press.
- Seyfort Ruegg, D. 2002. *Two prolegomena to madhyamaka philosophy*. Vienna: Wiener Studien zur Tibetologie und Buddhismuskunde.
- Siderits, M. 2003. *Personal identity and Buddhist philosophy: Empty persons*. Burlington, VT: Ashgate.
- Taylor, Charles. 1982. Rationality. In *Rationality and relativism*, eds. M. Hollis and S. Lukes. Cambridge, MA: MIT Press.
- Thanissaro Bhikkhu. 2006. The integrity of emptiness. In *Buddhadharma, the practitioner's quarterly*, Winter, Shambala, Boulder, CO.
- Tillemans, Tom. 1990. *Materials for the study of Āryadeva, Dharmapāla and Candrakīrti*. Vienna: Wiener Studien zur Tibetologie und Buddhismuskunde.
- Tillemans, Tom. 2004. What are Mādhyamikas Refuting? Śāntarakṣita, Kamalaśīla et alii on Superimpositions (*samāropa*). In *Three mountains and seven rivers, prof. Musashi Tachikawa's felicitation volume*, eds. S. Hino and T. Wada. Delhi: Motilal Banarsidass.
- Torella, Raffaele. 2007. Studies on Utpaladeva's Īśvarapratyabhijñāvivṛti. Part I: Anupalabdhi and apoha in a Śaiva garb. In *Expanding and merging horizons: Contributions to South Asian and cross-cultural studies in commemoration of Wilhelm Halbfass*, ed. K. Preisendanz. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- van Eemeren F.H., and R. Grootendorst. 2004. *A systematic theory of argumentation*. Cambridge: Cambridge University Press.
- Waldron, William S. 2003. *The Buddhist unconscious. The ālayavijñāna in the context of Indian Buddhist thought*. New York: Curzon.
- Wiggins, David. 1998. Weakness of will, commensurability and the objects of deliberation and desire. In *Needs, values, truth*, ed. D. Wiggins. Oxford: Oxford University Press.