Self-Conscious Emotions
Without a Self

Monima Chadha & Shaun Nichols
Monash University      Cornell University

Introduction
Certain metaphysical views are thought to have implications for the kinds of feelings that are appropriate to have. For instance, many philosophers maintain that we lack free will and that, as a result, reactive attitudes like resentment are inappropriate. Resentment would only be appropriate if people had genuine libertarian free will; since people lack such free will, we should not resent people even when they do us wrong (e.g., Pereboom 2001, Sommers 2007). Buddhist metaphysics also has implications for the kinds of reactive attitudes that are appropriate to have. Insofar as Buddhism denies the existence of a self, emotions that depend on a representation of self are based on a fundamental mistake.

In contemporary emotion theory, “self-conscious emotions” are those for which a representation of the self is implicated in the activation of the emotions. The most familiar examples of such emotions are pride, shame, and guilt. In Lazarus’ influential appraisal framework, he maintains that a critical component of the appraisal of pride registers that “credit is to oneself”. For guilt and shame, the appraisals must include that “blame is to oneself” (1991, 271; 242–3). More generally, Tracy and Robins note that “the primary distinctive characteristic of self-conscious emotions is that their elicitation requires the ability to form stable self-representations (‘me’), to focus attention on those representations (i.e., to self-reflect; ‘I’), and to put it all together to generate a self-evaluation” (Tracy and Robins 2007, 191).

As we will see, Buddhists abjure some emotions like pride and self-contempt precisely because they are self-conscious emotions. However, recent discussions of emotions in Buddhism suggest that one of the canonical self-conscious emotions, shame (the received translation of the Pāli term 'hiri'), is an emotion to be endorsed and indeed cultivated. The canonical texts in the Abhidharma Buddhist tradition, which provide the basis for our discussion in this paper, endorse hiri as one of the wholesome (kusala) factors “always found in all good minds” (Abhidharmakośabhāṣya, trans. Pruden 1988, Chap. 2.25, 190,) and as one of “the guardians of the world” (Atthasālīni, 124–5). Insofar as hiri is a self-conscious emotion, this is at odds with the central
Buddhist claim that we should rid ourselves of the idea that there is a self. Buddhist moral education seems to promote an emotion that fundamentally presupposes something that Buddhist metaphysics fundamentally rejects — a self. This puzzle provides the motivation for our paper, and we will argue for a new understanding of hiri that also has implications for how we should think about one important “self-conscious” moral emotion, guilt.

We will engage this philosophical puzzle posed by hiri on its own terms. But before we do that, we want to acknowledge two more general strategies that Buddhism might exploit to justify thoughts that invoke a self: The Doctrine of Two Truths and the Path. According to the Doctrine of Two Truths, while reference to a self cannot be ultimately true — since ultimately there is no self — a statement that refers to the self can be conventionally true provided acceptance of the statements reliably leads to successful worldly activities (Siderits 2008, 35). This practical element of Buddhist instruction is emphasized in accounts of the Path to enlightenment. The idea is that when one sets off on the Path prescribed by the Buddhists, one has no choice but to use conventional notions to make progress on the Path towards enlightenment. Thus, one might maintain that it is consistent for Buddhist philosophers to promote a self-conscious emotion (viz., shame) insofar as it is premised on a conventional truth or is an important factor on the Path to enlightenment. Although we acknowledge that advertence to the Two Truths or the Path can be gainfully deployed by the Buddhist in various contexts, we do not wish rely on these strategies here. The reason is that reference to the Two Truths or the Path can at most help explain why hiri (understood as a self-conscious emotion) is consistent with Buddhist practice, it cannot provide a theoretically satisfying account of the nature of the moral emotions. In particular, appeal to the Two Truths or the Path won’t help us determine which moral emotions are theoretically consistent with the no-self view.

The natural question that arises at this point then is: What is the right way to understand hiri? We will, of course, rely on Buddhist texts in exploring the puzzle of hiri. But our primary interest here is in the philosophical issues. The no-self doctrine is a central reason for the sustained philosophical interest in Buddhism, and this doctrine has implications for how to think about the propriety of various beliefs, practices, and emotions. So we want to explore whether hiri can be consistently retained as a positive emotion that ought to be cultivated if one takes the no-self view. As we’ll see, by taking the issue on its own terms (rather than advertence to the Two Truths or the Path), we can uncover a strand in Buddhism that solves the puzzle, revealing a notion of hiri as an important moral emotion that does not require representation of the self.

We begin with a discussion of Abhidharma model of mind as a background to the place and role of emotions in Abhidharma moral psychology. In Section 2, we present the Abhidharma typology of emotions and highlight the differences with Western accounts. We also discuss contemporary Buddhist accounts of hiri as a self-conscious emotion. Section 3 provides a new construal of hiri as anticipatory guilt, and we argue that anticipatory guilt doesn’t require an explicit representation of self. We close with some remarks about the benefits of cross-cultural philosophy in rethinking our classifications of emotion terms.

1. Abhidharma Model of Mind

In order to understand Buddhist treatments of emotions, we need to have in place a richer sense of Buddhist philosophy of mind and, in particular, the basis for their distinctions between different mental factors. The Buddhist analysis of experience maintains that what we experience as a temporally extended, uninterrupted flow of phenomena is, in fact, a rapidly occurring sequence of causally connected events, each with its particular discrete object. To explicate this, Buddhist philosophers decompose the world and ourselves in it into a causal sequence of evanescent mental and physical states (nāma-rūpa). Though there are various construals of this central Buddhist nāma-rūpa in the literature, the best way to understand this notion is that of a minded-body (Ganeri 2017, 77–9). In Abhidharma Buddhism, the minded-body
is further analyzed into “dharmas”, which are discrete momentary factors. Importantly, thus the dharmas include both physical and mental factors. On the Abhidharma picture, mind is not a substance or central processor that produces experiences and thoughts; rather it is an aggregate of many simultaneous series of mental dharmas. These mental dharmas are best thought of as “phenomenologically basic” features that constitute conscious experience (Dreyfus 2011). This does not, however, mean that the phenomenological features are readily available in ordinary introspection. The claim is that mental dharmas are in principle available in first-person experience, though discerning the dharmas requires meditation practice. Indeed, some dharmas are better thought of as subliminal mental factors that can be brought to the surface only through sustained meditation practice. The Abhidharma schools disagree about the number, classification, and role of these features in experience. So the Abhidharma philosophers take great pains to provide ever new lists and classifications of mental dharmas and detailed arguments to justify the proposed revision. However, Abhidharma schools agree on the starting point for grouping the mental factors: They are primarily classified as good (kusala), bad (akusala), and neutral (abyākata). Good (kusala) is defined as that which is “salutary, blameless, skillful” (Atthasālinī, 62–3) and thus reduces suffering. Bad (akusala) is just the opposite; it is unhelpful, blameworthy, unskillful and augments suffering. Some mental factors are wholesome or good in themselves, e.g., compassion, wisdom, etc.; others are unwholesome or bad in themselves, e.g., anger, greed, craving, etc.; and yet others are neutral e.g., equanimity, resolve, etc. The moral valence of a given conscious state or thought, i.e., whether it is good or bad, is determined by the moral valence of mental factors that comprise conscious thought and experience. For example, a thought associated with compassion would be good because compassion is a good factor; a thought associated with equanimity would be indifferent because equanimity is disinterested; a thought associated with greed and ignorance would be bad because greed and ignorance are bad factors.

The overarching aim of the Abhidharma philosophy is to cultivate the wholesome mental factors and eradicate the unwholesome ones. This, in turn, will ensure a prevalence of good thoughts, intentions, and actions, thereby reducing suffering. How does one go about identifying the good and the bad factors? The Abhidharma answer is to turn to the tradition as a repository of moral knowledge delineating the good and bad factors. However, experienced teachers also suggest a turn to moral phenomenology. The idea is to pay attention to how thoughts or actions appear or feel to a person. In developing their moral phenomenology, Buddhists begin by noticing that the pursuit of sense pleasures is typically mixed with hardship and disturbances in the mind because such pursuits involve greed and craving for more of the same. In contrast, by purifying the mind through restraining oneself from indulging in sense pleasures, one “experiences internally an unmixed ease (sukha)” (Majjhima Nikaya I 181). For example, by their very presence in the mind, loving-kindness and compassion have a calming influence and result in easing the mind. Good and bad thoughts and actions can both appear joyful and pleasurable, but only bad thoughts cause distress and disturb the mind. In the Abhidharma psychology, good or wholesome (kusala) thoughts are never painful or distressing, though they can be neutral. They are felt as neutral when they are experienced through equanimity and disinterest. The thought is that we focus on experientially available distinctions to figure out which mental factors are wholesome or good. Wholesome factors can be differentiated from unwholesome ones in that the former involve a healthy and uplifting state of mind in contrast to the latter that distress and disturb the mind.

Given the salient differences between the Abhidharma model of the mind and other models of mind in Western philosophy and psychology, it is natural to expect that the mental categories will be very different. The Abhidharma philosophers are not concerned with distinguishing emotions from other mental factors. They are, like contemporary emotion theorists, interested in the action-guiding role of particular emotions; but it bears emphasis that these actions are set in
the context of the guiding principle of reducing suffering in the world. Thus, the primary division among mental factors will be in terms of whether they reduce suffering (good) or increase suffering (bad) or have no effect on suffering (neutral). Ignoring the neutral factors for now, a partial list of good and bad factors is reflected in figure 1 to give the reader a sense of the Abhidharma typology.

<table>
<thead>
<tr>
<th>Mental factors with good moral valence</th>
<th>Mental factors with bad moral valence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right view (samma śatthi)</td>
<td>Wrong view (micchā satthi)</td>
</tr>
<tr>
<td>Right thought (samma sāṅkappa)</td>
<td>Wrong thought (micchā sāṅkappa)</td>
</tr>
<tr>
<td>Right effort (samma vāyāma)</td>
<td>Wrong effort (micchā vāyāma)</td>
</tr>
<tr>
<td>Right concentration (samma samādhi)</td>
<td>Wrong concentration (micchā samādhi)</td>
</tr>
<tr>
<td>Right mindfulness (samma samātī)</td>
<td></td>
</tr>
<tr>
<td>Non-greed (alobha)</td>
<td>Greed (lobha)</td>
</tr>
<tr>
<td>Non-hatred (adosa)</td>
<td>Hatred (dosa)</td>
</tr>
<tr>
<td>Non-delusion (amoha)</td>
<td>Delusion (moha)</td>
</tr>
<tr>
<td>Non-covetousness (anabhijjhā)</td>
<td>Covetousness (abhijjhā)</td>
</tr>
<tr>
<td>Non-malice (abyāpāda)</td>
<td>Malice (byāpāda)</td>
</tr>
<tr>
<td>Shame (hiri)</td>
<td>Shamelessness (ahirika)</td>
</tr>
<tr>
<td>Fear of wrongdoing (ottappa)</td>
<td>Fearlessness (anottappa)</td>
</tr>
<tr>
<td>Impartiality (tātramajjhata)</td>
<td>Pride (māna)</td>
</tr>
<tr>
<td>Compassion (karuṇā)</td>
<td>Self-contempt (omāna)</td>
</tr>
<tr>
<td>Sympathetic joy (muditā)</td>
<td>Envy (issā)</td>
</tr>
<tr>
<td></td>
<td>Avarice (macchariya)</td>
</tr>
<tr>
<td></td>
<td>Remorse (kukkucca)</td>
</tr>
</tbody>
</table>

Figure 1: Mental factors organized by moral valence

Given the guiding division in Buddhist typology between good and bad factors, we expect some differences in the types of emotions charted by the Buddhist Abhidharma schools and by contemporary emotion theorists. Consider, for instance, contemporary evolutionary psychological accounts of the emotions. Those views focus on how emotions are adaptations to ancestral problems. As a result, researchers in this tradition do not typically distinguish between beneficial and harmful emotions (Cosmides and Tooby 2000). Even those who categorise emotions as “positive” or “negative” often do not regard “negative” emotions as always harmful to oneself or to others (Watson, Clark, and Tellegen 1988). Rather, in concert with Aristotle’s model, emotion theorists often maintain that all emotions are conducive to one’s well-being, as long as they are not excessive or inappropriate to the context. Thus, in this tradition, there is considerable focus on emotion regulation where the idea is not to rid oneself of the emotion altogether, but to cope with excessive or inappropriate instances of the emotion (Davidson, Jackson, and Kalin 2000). Another important difference between the Abhidharma and contemporary emotion theorists is that, according to the Abhidharma, there is a match between the felt valence and the positive or negative force of the emotion. According to Abhidharma philosophers, the felt valence of the particular emotions, the “feel calm” factor that enhances our peace of mind or “internal disturbance” that agitates our mind is a reliable guide to whether those emotions are positive or negative respectively. By contrast, contemporary emotion theorists think of some emotions as having a positive function even though the peaceful and calming feeling alleged by Abhidharma philosophers is not associated with those specific emotions. Guilt is one such example — although it feels painful, it is thought to have positive function insofar as it discourages bad behaviour and plays an important role in reparation of loving relationships (see, e.g., Frank 1988, Baumeister et al. 1994). One can feel quite agitated and disturbed by guilt, but many emotion theorists celebrate it as motivating morally commendable (and personally beneficial) behavior. More generally, on contemporary accounts of emotions in psychology, there is no presumption that there will be a match between felt valence and the negative or positive contribution of the emotions.

It is obvious then that the Abhidharma way of cutting the pie of what we call the mind and its emotions is different from the Western typologies of the mind and its emotions. And these apparent
differences have important implications for how we resolve the puzzle identified in the Introduction. In important ways, the Buddhist typology of emotions is more fine-grained than the Western typologies, and we will draw on the more fine-grained Buddhist typology to develop our account of ātītāna.

In this section, we have articulated how the distinction between harmful and beneficial emotions is central to Buddhist emotion typology. Emotions on this Buddhist picture are either to be cultivated or eradicated, so they cannot be both harmful and beneficial. In the next section, we shall see that we also need to distinguish between forward-looking or anticipatory and backward-looking or retrospective emotions. As we will see, this parsing of emotions typology across traditions provides a natural way to dispel the apparent inconsistency in the Abhidharma Buddhist treatment of self-conscious emotions; in particular, the central emotion ātītāna need not have the problematic invocation of a self. We now turn to standard treatments of self-conscious emotions in the contemporary Buddhist literature to show why this issue needs to be reconsidered.

2. Abhidharma on Emotions

The Abhidharma are not especially concerned with distinguishing emotions from other mental factors. Instead, the Buddhist Abhidharma focus on identifying and classifying types of mental activity into those that are conducive to reducing suffering, and others that are toxins in that they tend to increase suffering. This classification is determined by whether they have an uplifting or a disturbing effect on the mind. Thus, it is not a surprise that the Abhidharma classifications place emotions like hatred and regret as unwholesome mental factors and compassion and lack of greed as wholesome mental factors.

The Ābhāsidharmika explicitly discourage self-conscious emotions like pride and self-contempt. Both pride (atimāna) and self-contempt (omāna) are derived from the root māna, translated as conceit. Conceit is defined as “...fancying (deeming, vain imagining). It has haughtiness as characteristic, self-praise as function, desire to (advertise self like) a banner as manifestation, greed dissociated from opinionatedness as proximate cause, and should be regarded as (a form of) lunacy” (Atthaśāliṇī, 256). In this sense, conceit is taken to implicate a self-representation, so too is its manifestation as in pride and opposite of pride as in self-contempt. And insofar as both pride and self-contempt are self-conscious emotions, they are seen as depending on a wrong view of the self. The Abhidharma denounce pride and self-contempt precisely because they require representing oneself as being superior or inferior to others respectively. Indeed, the point here is that all kinds of self-conception and representation must be “penetrated” (understood as having no basis in reality) and rejected, as they can only increase suffering. As we saw, in the Abhidharma corpus, moral value is essentially tied to reduction of suffering. Dharmas or mental factors are classified as bad when they augment suffering, e.g. greed has negative moral value because it produces craving (tanha). Craving is the key factor in producing suffering. Thoughts and intentions based on craving will increase suffering, and insofar as such thoughts are the basis of one’s desires and wants and actions, they all have negative moral value. Pride and self-contempt have negative moral value, too, because they are based on a false self-conception that is causally implicated in producing craving for existence (bhava-tanha), which in turn leads to increasing suffering.

The Buddhists regard false self-conception as the fundamental basis of craving for existence and suffering. As a result, Buddhism encourages us to extirpate thoughts like I will exist in the future or I want more. But this doesn’t mean that everything that gets called “self-reference” needs to be rejected. Take the predicates “heavy” and “to the left” as used in sentences like “The book is heavy” and “The ball is to the left”. These predicates are implicitly relativized to an individual, but this kind of “self-reference” does not implicate a problematic kind of invocation of self. Predicates like “heavy” and “to the left” can be indexed to the body, and the body is not denied by Abhidharma Buddhism. On the Abhidharma view, bodies exist — they are constituted by aggregates of momentary material dharmas. Thus, the self that...
is denied by the Ābhidharmika need not be invoked to explain the “self-reference” of “heavy” or “to the left”. By contrast, self-conscious emotions like pride and self-contempt seem explicitly to implicate a notion of the self as enduring and self-aware, and that kind of self-reference is exactly what Buddhism seeks to extirpate. Thus, it makes sense that Buddhists would repudiate such self-conscious emotions.

We now turn to the puzzling case of hiri. In the most detailed discussion of self-conscious emotions in contemporary Buddhist scholarship, Maria Heim (2009, 244–5) notes that the Abhidharma literature does not seem to have an equivalent for guilt. Hiri is a key positive emotion for Buddhists, and it finds a parallel in ottappa, typically translated as fear of wrongdoing. Heim distinguishes guilt from hiri in two ways. First, she says that hiri depends on “social norms and values developed through seeing oneself through the gaze of another” as opposed to guilt which is “an internally generated anguish of having committed a wrong”. Heim translates hiri as shame. We will suggest an alternative way of thinking about hiri on which hiri is akin to a kind of guilt.

But we want to emphasize here Heim’s second observation about guilt. She notes that unlike guilt, both hiri and ottappa are forward-looking. The value of hiri and ottappa “lies in what they keep us from doing, not in wretched anguish when reflecting on wrongs already committed” (Heim 2009, 245). Heim is here drawing on a key feature of Buddhist typology. As we saw in Section 1, Buddhist typology is driven by whether a mental factor is good or bad, including especially whether a mental factor helps to increase or reduce suffering. As a result of this taxonomic principle, insofar as backward-looking mental factors won’t help reduce suffering, those factors will be systematically excluded from the set of good mental factors. Contemporary emotion theory treats guilt as a single emotion that can be generated from recalling past misdeeds or from contemplating future misdeeds. This fusion of backward- and forward-looking factors is rejected by the Buddhist typology. Thus, insofar as guilt is taken to be a single emotion category that encompasses both retrospective and prospective responses, guilt is absent from the Buddhist typology.

Given Buddhist typology, there is nothing corresponding to the unified notion of guilt found in Western typologies. But the Abhidharma literature does have a term, kukkucca—translated as regret or remorse—which is akin to the backward-looking aspect of guilt. As one would expect, the Abhidharma associate kukkucca with agitation, which increases suffering, and thus kukkucca is found in the group of bad or unwholesome factors. Heim notes that kukkucca is explicitly discouraged in Pāli sources. She quotes from Buddhaghosa’s Atthasāliṁī, a key text in the Abhidharma tradition: “Since one cannot undo a bad deed nor do a good deed that was neglected, returning again to it in remorse is ugly” (Heim 2009, 246). The Abhidharma Buddhist is quite clear on this, maintaining that one who is enlightened (an arhat) is not troubled by remorse or worry about evil deeds done in the past.

Although pride and remorse are discouraged by Buddhists, hiri is recommended. As noted above, the Abhidharma sources present hiri paired with ottappa (fear or apprehension of wrongdoing) as wholesome mental factors, considered to be the “guardians of the world” (Anguttara Nikaya 2.9). So, what is hiri? Heim claims that hiri is best translated as shame, since hiri “can be nonmoral and refer to modesty” but it can also “have moral connotations tied up with self-respect” (2009, 244). We will question this interpretation of hiri later in this section. First, though, we want to set out Heim’s account a bit more fully.

Hiri and ottappa are considered morally valuable in that they prevent us from doing harmful actions. One distinctive feature of hiri, according to Heim, is that it is primarily a future-oriented state, morally valuable for what it offers at the moment of decision-making rather than concerning the deeds done in the past. This feature is emphasized by pairing hiri with ottappa rather than regret (which concerns the past); together, hiri and ottappa are listed as beautiful or wholesome mental states—as powers and guardians of the world.1 Buddhaghosa defines them thus:

1. Heim emphasizes that Buddhaghosa, in line with the tradition, regards hiri as a beautiful or wholesome factor. But she interprets hiri as shame, which is typically associated with aversive feelings. If hiri is aversive, this seems
Hiri has the function of not causing evil because of the condition of shame, ottappa because of the condition of fear. They are manifested in shrinking from evil in the manner already described; their proximate causes are respect for self and respect for others. Having respect for oneself, one resists evil because of hiri, like the daughter of a good family. Having respect for others means one resists evil because of fear, like a prostitute. And so these two states should be regarded as guardians of the world (Visudhimagga 464–5, cf. Atthasālinī, 124–5).

The thought here is that hiri arises from self-respect, suggesting that hiri involves the anticipated failure to live up to a standard that one sets for oneself. The Atthasālinī notes that the internalized standards for hiri conform to one's social standing in the community, one's age, and learning in deference to social norms (Atthasālinī, 127). Although there is talk about conformity with social norms, and these norms are sensitive to one's class, age, and learning, nonetheless, the Pāli literature is unequivocal that hiri does not imply conformity to any old social norms — the relevant norms are tied to the behavior of admired monks, teachers, the Buddha himself, and idealized others in the larger monastic community, the sangha. A monk's self-appraisal is tied to seeing himself or herself as part of the sangha.

In contrast to hiri, ottappa is supposed to be generated out of fear of external reprisal. It is described as a visceral fear from anticipating not living up to external standards set by one's community. Buddhaghosa illustrates this in his remark that ottappa allows us to resist evil like a prostitute while hiri allows us to resist evil like the daughter of a good family. Prostitution in ancient India was a legal profession, somewhat like a contract between the prostitute and her client under the jurisdiction of the state. The prostitute would not cheat (her client) because she could be fined or punished for breaking the contract. A daughter of a good family would not cheat (anyone) because she had internalized the fact that cheating was wrong, even if there was no reprisal. On a related note, Heim notes that while ottappa is generated by external standards that one is expected to live up to (lokadhīpatti or "to be ruled by the world"), hiri is generated by internal standards one sets for oneself (atṭadihīpatti or "to be ruled by the self"). Again, we see this reflected in the Atthasālinī:

How is ottappa understood as ruled by the world? Here a certain son of a noble family places the world foremost as his ruler and does not do a harmful act (Atthasālinī, 126).

How is hiri governed by the self? Take the son of a noble family, who makes self as his chief influence, and so refrains from evil thinking, ‘It is not fit that such a man as I, who left the world through faith, endowed with wide experience, believing in the ascetic life, should do evil.' Thus hiri is governed by the self (Atthasālinī, 148).

In arguing that hiri is best construed as shame, Heim draws attention to the fact that in the Atthasālinī, Buddhaghosa says that hiri is the appropriate reaction to realizing that one is seen by a respected elder while relieving oneself (Heim 2009, 243). As Heim notes, this is similar to Bernard Williams’ suggestion that the most primitive experiences of shame are connected with sight and being seen (Heim 2009, 240–1). In shame, one wishes to disappear, “to sink through the floor” (Williams 1993, 89). It is interesting to note that Williams also thinks of shame as a positive emotion that should be inculcated. A primitive sense of shame, Williams suggests, is an incentive to achieve human

---

Footnote:

To be at odds with Abhidharma moral phenomenology, according to which aversive feelings indicate a negative valence, rather than a wholesome factor. To be sure, this concern will also apply to our suggestion that hiri is rightly interpreted as anticipatory guilt. Insofar as shame and guilt elicit aversive feelings, this seems to conflict with the Abhidharma on account of the moral phenomenology of wholesome factors. This is a large issue that we cannot take up adequately here, but one conciliatory possibility is that while an occurrence of hiri might be partly aversive, hiri will also deter harmful actions and that generates an overall positive feeling. We will not pursue this point in this paper as it is orthogonal to our central point that hiri does not require representation of the self.

---
excellence. This thought has a long and venerable history in Western philosophy. In the *Laws*, Plato claims that shame is what will protect man from doing what is dishonorable (647a). Williams adds that acting out of shame need not mean acting out of a childish concern with how one appears to others. There is a distinction to be made between a childish desire for the approval of others and a mature desire to be respected by those whom one respects and whose ethical sensibilities one has internalized as a part of developing one’s moral character.

The similarities between Williams’ account of shame and Buddhaghosa’s account of *hiri* are noteworthy, but are they the same concept? To answer this question, it is worth asking whether the notion of “self” in phrases like ‘respect for self’ and ‘being governed by the self’ invokes the false self-conception in a way that is problematic for Abhidharma philosophy? Or does it just invoke a more innocent reference to self as a body or a minded-body (*nāma-rūpa*)? The minded-body, on the Abhidharma construal, is nothing over and above evanescent mental and physical factors; it does not invoke the problematic notion of a self-same entity enduring through time. The notion of self implicated in these expressions is consistent with minded-body since in comparing *hiri* and *ottappa*, Buddhaghosa is drawing attention to a contrast between standards that are internalized (as in standards one sets for oneself) and those that are external (as in standards that one owes for fear of reprisal from society). The internalized standards explain the example of constraints on the behavior of the “daughter of a good family”. A minded-body can easily provide the basis of such a contrast between internal and external standards. There is no need to invoke the false self-conception, insofar as *hiri* does not involve an essential reference to an enduring self. There is at least some reason to doubt Heim’s translation. We return to this point in the next section.

Most contemporary philosophers think that shame has some positive aspects, and are not willing to treat it as totally negative (Williams 1993; Velleman 2001; Deonna et al. 2011). Many contemporary psychologists, in contrast, regard shame as a negative emotion, in that it involves a negative assessment of a *stable global self* and leads to escapist and hiding behaviors. A recent paper distinguishes between guilt and shame in the following way:

... guilt is concerned with the negative evaluation of a specific behavior (‘I did that wrong’) resulting in a desire to confess, apologize, and repair, shame pertains to the negative evaluation of the self (‘I did that wrong’) causing a desire to vanish, escape, or strike back...guilt should be primarily viewed as a ‘good’ emotion; because of its positive associations with morality and empathy, this self-conscious emotion probably prevents the development of externalizing (i.e., disruptive behavior) problems. In contrast ... shame as ‘bad and ugly’, as excessively high levels of this self-conscious emotion likely promote inferiority, self-punishment, and defensive aggression, and in its wake would make individuals more susceptible for developing both internalizing (i.e., emotional) and externalizing problems (Muris et al. 2018, 268; see also Tangney and Dearing 2002).

These analyses of shame seem to imply that self-appraisal and self-evaluation are part and parcel of the notion of shame insofar as it is a self-conscious emotion. Heim is quite aware of this. In describing the contrast between shame and guilt, while quoting Williams she writes, “Guilt looks to the wrong committed, or its victim, while ‘shame looks to what I am’” (2009, 248). To be sure, Williams celebrates shame and Buddhists celebrate *hiri*. But it is surprising that Heim fails to note or comment on the fact that the invocation of the self in these analyses is diametrically opposed to the Abhidharma view that there is no-self. In the same vein, Jonardon Ganeri (2017, 317) suggests that *hiri* is akin to the concept of conscience as it involves the general exercise of a capacity for self-evaluation. But self-evaluation presupposes a self as the object of evaluation.

Shame as described by Western philosophers and psychologists is far from what we might have antecedently expected the Abhidharma
philosophers to mean by hiri. Hiri is clearly regarded as a wholesome basic mental factor, one which should be cultivated rather than eradicated. Hiri has positive valence and is beneficial insofar as it prevents us from doing evil. However, if shame, like pride, requires a false self-conception of the self as the self-same entity enduring through time, then it is certainly ego-involving — caught up with the conceit of “I am”. This would seem to entail that shame is unwholesome since it has its roots in a cognitive error or wrong view: a false presupposition of a self. Indeed, this is particularly striking since hiri is regarded as a dharma; if hiri really were ego-involving it would be the only dharma which is ego-involving and wholesome at the same time. There are other ego-involving dharmas (e.g., anger, conceit), but they are unwholesome, and they are to be eradicated together with the expiration of the sense of self. Apart from the puzzling case of hiri, all other emotions that count as wholesome dharmas (e.g., compassion and sympathetic joy) are not ego-involving and can be cultivated consistently with the eradication of the sense of self.

On behalf of the Abhidharma philosophers, we will argue for a different way of thinking about hiri. Shame is a self-conscious emotion entailing that the object of evaluation is an enduring self. But there is no enduring self for Abhidharma, so what is the object of evaluation? Our solution to this puzzle will be to treat hiri as a kind of anticipatory guilt rather than as shame.

3. Reconceptualizing Hiri

As we’ve seen, the standard treatments of guilt and shame in the contemporary literature on the Abhidharma account of emotions (Heim 2009) suggests that guilt is not even recognized in the Buddhist typology of emotion and that shame is a vitally important emotion for Buddhist practice. We will argue that once we recognize some distinctions, we can see a more promising alternative. To explicate this, we want to review standard accounts of guilt and shame in contemporary emotion theory. Emotion systems are standardly characterized in terms of three elements: (1) the function of the emotion, (2) the inputs that activate the emotion, and (3) the outputs of the emotion, including prominently the behavioral outputs or “action tendencies”.

In the case of guilt, the function is supposed to include the repairation of relationships after doing harm to a loved one (Baumeister et al., 1994). In addition, guilt has a forward-looking function — it serves to guard against doing harmful things in the first place. In the language favored by economists, guilt makes us less likely to cheat in cooperative enterprises. Robert Frank writes:

Consider, for example, a person capable of strong guilt feelings. This person will not cheat even when it is in her material interests to do so. The reason is not that she fears getting caught but that she simply does not want to cheat. Her aversion to feelings of guilt effectively alters the payoffs she faces (1988, 53).

Frank maintains that guilt is an evolutionary adaptation (see also Trivers 1971). But the basic point here about the function of guilt is also adduced by those who maintain that guilt is acquired through social learning (e.g., Prinz 2004, 126–7). The core idea is that the anticipation of guilt serves the important function of discouraging harmful behavior. So in addition to the function of repairing the damage one has done for past actions, guilt has the function of discouraging future harmful behavior through the anticipation of aversive guilt feelings (see also Fessler and Haley 2003, 16).

There is less consensus about the function of shame. But a plausible gloss is that shame functions to encourage social conformity (e.g., Fessler & Haley 2003, 20). As with guilt, there will be backward and forward variants. When one has done something that caused shame (typically by breaking a social or moral norm), shame will serve to restore or salvage one’s standing in the community, and this function is served by withdrawal or efforts at appeasement (see, e.g., Tangney & Dearing 2002, Gilbert 1997). When one considers violating a social or moral norm in the future, the anticipation of shame would discourage the behavior.
Thus, the function of guilt is to discourage harmful actions and to repair damage caused by past harmful actions. Shame’s function seems to be somewhat different. To be sure, shame discourages harmful actions. But shame also seems to have the function of discouraging non-moral social mistakes. For instance, children often feel shame for wetting their pants and part of shame’s function is plausibly to ameliorate or avoid such social faults.

Now let’s consider the outputs. If a person feels guilty for a past action, the action tendency will be to repair the relationship with the victim. If a person feels guilt in anticipation of a harmful action, the output will first be an aversion to the action, which will presumably lead to fewer instances doing harmful actions.

For the case of shame, if a person feels shame about a past action, the action tendency is to appease or withdraw (Izard 1977, Gilbert 1997, Fessler 1999, Lazarus 1991, 244).

Fessler & Haley write: “Appeasement serves to maintain the relationship despite the disparity, while avoidance serves to minimize exploitation” (2003, 18). The output of anticipatory shame will be similar to the output of anticipatory guilt — when considering some potential course and shame is anticipated, one will be less inclined to do the action.

Finally, let’s turn to the inputs that trigger the emotions. Here, emotion theorists have argued for a central difference between shame and guilt, along the lines suggested in Section 2. A person can feel either shame or guilt when reflecting on a past moral transgression or when contemplating a harmful action (Tangney and Dearing 2002, 17–8). So what is the difference between the triggers for shame and guilt? It is precisely the role of the self in the appraisal. In a widely influential treatment, Helen Block Lewis argues that shame attaches to the self and guilt to the action. She writes:

The experience of shame is directly about the self, which is the focus of evaluation. In guilt, the self is not the central object of negative evaluation, but rather the thing done or undone is the focus. In guilt, the self is negatively evaluated in connection with something but is not itself the focus of the experience (1971, 30; Tangney & Dearing 2002).

So I feel shame when I think “I did that (bad thing)” and I feel guilt when I think “I did that (bad thing)”. Similarly, when I anticipate doing a bad thing, if my focus is on my self doing the thing, shame is the likely result, whereas if I focus on the act itself, guilt is the likely result. (See Figure 2 for a summary of the differences between shame and guilt.)

<table>
<thead>
<tr>
<th>Function</th>
<th>Guilt</th>
<th>Shame</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inputs</td>
<td>Recognition that I did that bad thing</td>
<td>Recognition that I did that bad thing</td>
</tr>
<tr>
<td></td>
<td>contemplating doing something harmful</td>
<td>contemplating being the source of something harmful</td>
</tr>
<tr>
<td>Outputs</td>
<td>Reparative behavior; deter harmful behavior</td>
<td>Withdrawal/appeasement; deter moral and social violations</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Profiles of guilt and shame

Before we turn to hiri, we want to recruit another distinction. As we saw in Section 2, Buddhists distinguish between anticipatory and retrospective emotions, and this is at least partly because they hold that retrospective emotions don’t help to reduce suffering. When we apply this distinction to shame and guilt, instead of two emotions, we get four: retrospective guilt and shame and anticipatory guilt and shame. In her discussion of hiri, Heim maintains that guilt always involves the past (2009, 249). But as we’ve seen, emotion theorists certainly accord a role for forward-looking guilt (Frank 1988, Fessler and Haley 2003, Prinz and Nichols 2010). Nonetheless, emotion theorists have not generally used retrospective and anticipatory factors in their taxonomies.
of emotion, and guilt is treated as a unified emotion across its retrospective and anticipatory versions (see, e.g., Haidt 2003, Fessler and Haley 2003, Prinz and Nichols 2010). However, when we look at the different elements of guilt and shame, we find obvious differences in function, output, and input, depending on whether the emotion is anticipatory or retrospective. To be sure, the function and outputs of guilt always concern harming others, but the differences between anticipatory and retrospective guilt are also quite salient: Retrospective guilt motivates repair, anticipatory guilt deters harmful action. This way of thinking about guilt should make us concerned about treating it as a self-conscious emotion as is standard in contemporary emotion theory. It’s also likely that the phenomenology of these emotions is quite different. Since we know that hiri is a forward-looking emotion, when we try to glean what hiri is, it seems that we need to focus on the anticipatory versions of guilt and shame, summarized in Figure 3.

<table>
<thead>
<tr>
<th>Function</th>
<th>Anticipatory guilt</th>
<th>Anticipatory shame</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outputs</td>
<td>Discourage harmful behavior</td>
<td>Discourage moral and social violations</td>
</tr>
<tr>
<td>Inputs</td>
<td>Deter harmful behavior</td>
<td>Deter moral and social violations</td>
</tr>
<tr>
<td></td>
<td>Contemplating doing something harmful</td>
<td>Contemplating being the source of something harmful</td>
</tr>
</tbody>
</table>

Figure 3: Profiles of anticipatory guilt and anticipatory shame

Now that we have the distinction between anticipatory guilt and shame in place, what is the best interpretation? First, what is the function of hiri? We find a very direct answer to this in Buddhaghosa: “Hiri has the function of not causing evil” (Visuddhimagga, 464–5; quoted in Heim, 238). This seems much closer to the function of guilt than shame, as shame has the broader function of not causing evil or social improprieties. What is the output of this future-directed emotion? Buddhaghosa says that it is “shrinking from evil”, which again is the distinctive output of anticipatory guilt, not the more general aversion to social wrongs characteristic of shame (Visuddhimagga, 142).

Although Buddhaghosa’s explication of the function and outputs of hiri fit with anticipatory guilt, in the very same passage he also says that the proximate cause of hiri is “respect for self”. In addition, elsewhere Buddhaghosa explicitly includes examples of hiri that really do seem to involve social improprieties: “For whenever the body is revealed, hiri is disturbed, damaged; thus it is called private parts, because of the disturbance of shame” (Visuddhimagga, 31). Heim also refers to Atthasālinī and Jataka tales where hiri is conceptualized as a sense of modesty or embarrassment, as in being seen while relieving oneself (Heim 2009, 243). Being embarrassed by nudity or bodily functions is characteristic of shame, but not moral shame. Our interest here is in the notion of hiri as a moral emotion.

We think that part of the explanation for the apparent inconsistencies here is that the Abhidharma corpus, like the Buddhist tradition, is large and diverse, and different senses of hiri might be intended in different places. Indeed, this point is already made by Vasubandhu, who is the most influential Abhidharma thinker, well-known for producing the seminal texts of at least three distinct scholastic traditions: Vaibhāṣika, Sautrāntika, and Yogācāra. In his *Treasury of Metaphysics* and its commentary (*Abhidharmakośabhāṣya*), Vasubandhu is careful to note that hṛi (Sanskrit for hiri) is of various types. He has in mind here the non-moral uses of hṛi as modesty, and as we saw above, Buddhaghosa sometimes uses hiri in this way. But Vasubandhu is very clear that insofar as he is talking about hṛi as a dharma, the simple definition is “Respect is Hṛi”. Other equivalents he proposes are “respect, veneration, fearful submission,” none of which invoke the self; they can instead be conceived of as respect for internalized standards (*Abhidharmakośabhāṣya*, Chap. 2, 32c, translated by Pruden 1988, 201). Indeed, it is worth noting that in the discussion of hṛi as a dharma in *Abhidharmakośabhāṣya*, there is no mention of self or oneself. In some sense, this should not be surprising as Vasubandhu would have had clearly in mind the importance of not adverting to a self. Nor is this
treatment of *hiri* unique to Abhidharma. The earliest mention of *hiri* in the Buddhist sources is in the *Advice to Rāhula at Ambalāṭṭhika*, which is taken to be the Buddha's advice to his son Rāhula. In this discussion, *hiri* is described in the context of evaluating *actions* rather than *self*. Using the example of telling a lie, Buddha emphasized the need for Rāhula to develop a sense of *hiri* around his actions: “Rāhula, do you see this hollow, empty water vessel?” — ‘Yes, venerable sir.’ — ‘Even so hollow and empty, Rāhula, is the asceticism of those who are not ashamed to tell a deliberate lie. Rāhula, when one is not ashamed to tell a deliberate lie, there is no evil, I say, that one would not do. Therefore, Rāhula, you should train thus: “I will not utter a falsehood even as a joke”’ (Majjhima Nikaya, 61). The passage goes on to say in contemplating whether or not to do a certain action by the body, speech, or mind, one should always reflect on it to check whether it leads to any harm to oneself or another, and if it does, one should avoid doing it on account of *hiri*. Insofar as *hiri* is used in the context of evaluation of *actions* in the *Nikayas*, there is reason to doubt that *hiri* is shame. Given that the tradition doesn’t explicitly distinguish between anticipatory shame and anticipatory guilt, it should not be surprising that *hiri* is not systematically treated as anticipatory shame. And once we have the distinction between anticipatory and retrospective emotions, it becomes plausible that *hiri* is often best interpreted as anticipatory guilt rather than anticipatory shame, as in Vasubandhu and in the above passage from the *Nikayas*. Of course, the standard account of guilt treats it as a self-conscious emotion, and thus, it seems like we are back into the same kinds of problems that we raised for shame. It is plausible that when we consider past actions, we only feel guilt when we think we were the agents of the bad action. Even if the focus of guilt is on the act that was committed rather than the self, it is still the case that we don’t feel guilty for an action if we find out that someone else did it. The situation with anticipatory guilt might be interestingly different, however. Recall that on the standard view, the input for shame is “I did that (bad thing)” and the input for guilt is “I did *that* (bad thing)”. In its anticipatory mode, shame would still foreground the self

— one contemplates oneself doing the bad or embarrassing thing. But for guilt, the focus is on the *action*, not self. In general, when we contemplate various options in decision-making, we need not represent the self in the bargain. So if I’m trying to decide which apartment to rent, I might have thoughts along the lines of “this room has a lot of natural light” or “that is a nice stove” or “this place is super expensive”. Those are the kinds of thoughts that are necessary for making decisions given one’s interests. But one need not also think “I like natural light” or “I prefer gas stoves”. The self-representation there is otiose. Similarly, when a person contemplates a harmful action, she might just think “lie?” rather than “do I lie?”. Insofar as she finds lying morally problematic, contemplating lying might trigger guilt feelings even if she doesn’t explicitly represent her self.

We would like to explain a bit more fully how it’s possible that anticipatory guilt could be triggered without a representation of self. To begin, we want to illustrate how self-representations aren’t necessary for decision-making. Production systems have been an important approach to modeling decision-making for decades. A simple production system is composed of a set of if-then statements, where the antecedent specifies a condition and the consequent specifies an action. If the antecedent is registered as satisfied (e.g., because it is represented in a working memory system), the program will generate the act specified in the consequent. Thermostats provide a familiar illustration (e.g., Newell 1990). A thermostat might have the following production system:

If temperature is greater than 66 degrees and heater is on, turn heater off.

If temperature is less than 77 degrees and AC is on, turn AC off.

A thermometer provides the information for the antecedent in this case, e.g., by generating a representation of ambient temperature. Notice that this system doesn’t contain anything like a self-representation.
It doesn’t say whose temperature is being registered or which heater is being turned on. Why doesn’t it need to specify my temperature or my heater? Because the system only measures the ambient temperature and it only controls one heater.

Now, turn to a simple case of decision-making. Imagine a vegetarian, Hannah, trying to decide what to eat at an Indian restaurant. She might think, “All the tandoori dishes have meat in them. The vindaloo is probably really spicy here. The saag sounds good…” None of this requires self-representation. Hannah doesn’t have to think, “The tandoori dishes have meat and I’m a vegetarian”, neither does she have to think, “The vindaloo is spicy and I don’t like spicy food”. In her decision-making, Hannah’s preferences guide her decision, but they don’t need to be represented as belonging to her. Just as with the thermostat, Hannah’s decision-making system doesn’t have to specify that the ethical views and tastes that are guiding the decision are hers. To be sure, one can register the difference between an endogenously generated action and an exogenously produced motion of the body. But this need not require that in the endogenous case, one represents the self as generating action. Instead, action might be the unmarked case of behavior production and what is noticed is when external forces compel behavior (Chadha 2017, 197; see also Prinz 2012).

So far, we have tried to explain how it’s possible for decision making to transpire without self-representations. There is even some incidental evidence that self-representation is absent from decision-making in many cases. The evidence comes from studies in which people “think aloud” while solving a problem — that is, they report what is running through their mind as they are solving a problem. This work in “concurrent protocol analysis” shows that people are quite accurate at reporting how they solve problems. For instance, when people are asked to solve a simple arithmetic problem like 19*4 and “think aloud” while they do it, people are excellent at reporting the process. The accuracy of self-report is confirmed by independent measures (see Ericsson and Simon 1993). Protocol analysis plausibly gives an accurate window into at least some critical aspects of cognitive phenomenology.

When we turn to the issue at hand, the representation of the self in decision-making, we find that when people think aloud during judgment and decision tasks, they rarely mention the self (except incidentally — “I guess”, “I would say”) (see, e.g., Horgan & Nichols 2016, 164–5). In hindsight, this isn’t very surprising. There is typically no need to think explicitly about yourself when you’re trying to figure out what to have for lunch or what to listen to or what to read. Your preferences guide those decisions with no need to call up the idea that these preferences belong to your self.

We noted above that sensible and efficient cognitive architectures like production systems don’t need a self-representation for basic thought and action. And the evidence from protocol analysis suggests that during decision-making, the self isn’t typically represented. This point plausibly carries over to the case of anticipatory guilt. Anticipatory guilt is a response to considering a harmful course of action, typically when engaged in decision-making. When we are trying to decide what to do, with various options before us, the contemplation of the harmful options can trigger anticipatory guilt. If I am tempted to lie to a close relative about not being able to visit, I feel anticipatory guilt and that makes me less likely to lie. What am I responding to here? It seems like I’m responding to the disrespectfulness and harmfulness of lying to a loved one — I’m responding to the moral properties of the action. But as with decision making in general, one doesn’t need to represent the self when considering immoral options. When considering a harmful act like lying, one need not think, “Do I lie?” or “I disapprove of harming loved ones” any more than Hannah needs to think, “Do I order meat?” or “I dislike spicy food” when deciding on dinner. Rather, just as one can respond to the gustatory properties of a dish without representing the self, one can respond to the moral properties of a prospective action without representing the self.

If this is right, then the input conditions for anticipatory guilt can be compatible with a no-self view. There are several significant lessons here. By following the Buddhist distinction between anticipatory and retrospective emotions, we see that there is an interesting
and underappreciated difference between two kinds of guilt. Western emotion theory treats guilt as a unified emotion and so treats guilt in general as a self-conscious emotion. However, once we prize apart the retrospective and anticipatory notions, we can see that anticipatory guilt can be triggered without any representation of self. In addition to this general lesson for emotion theory, we can give an account of the Buddhist view of the wholesome emotion *hiri*. If we take *hiri* to be anticipatory guilt, this fits well with the texts that describe *hiri* as “respect” and as having the function of preventing evil deeds. And given that anticipatory guilt doesn’t require a self-representation, on this reading, promoting *hiri* is consistent with a thoroughgoing no-self view.

**Conclusion**

We started with a puzzle. How can Buddhist philosophers who deny the self promote an emotion that requires thinking there is a self? *Hiri*, one of the “guardians of the world”, is standardly treated as shame, an emotion that is taken to require thinking about and evaluating the self. This question about the Buddhist tradition also raises a basic philosophical question: What kinds of moral emotions are theoretically consistent with the denial of a self? We have argued that anticipatory guilt might be such an emotion, and that it provides a plausible interpretation of *hiri* in key Buddhist texts.

Our argument is an instance of cross-cultural philosophy. The revisionist metaphysics and distinctive moral phenomenology advocated by Abhidharma philosophers dictates a unique typology of mental states, focused on whether the states are positive or negative with respect to reducing suffering. The positive and negative typology of Abhidharma Buddhism leads Buddhists to emphasize the distinction between anticipatory and retrospective emotions; in contemporary emotion theory, this distinction doesn’t play a typological role. At the same time, the Buddhist tradition largely elides the distinction between guilt and shame, which is prominent in contemporary emotion theory. When we apply both the Abhidharma Buddhist distinction and the contemporary distinction, we arrive at a new set of emotion categories: retrospective guilt, retrospective shame, anticipatory guilt, and anticipatory shame. Moreover, insofar as anticipatory guilt involves contemplating various acts (e.g., *lie* or *tell the truth*), it need not involve self-representation in the way shame and pride do. In addition, this notion of anticipatory guilt provides a plausible interpretation of *hiri* in several important Buddhist texts. This dispels the threat of incoherence faced by Abhidharma metaphysics and moral phenomenology. It also indicates that in its anticipatory form, the paradigmatic “self-conscious” emotion of guilt need not be self-conscious at all.2

**References**


2. We are very grateful to Georges Dreyfus and anonymous reviewers for Philosophers’ Imprint for useful comments on previous drafts that helped us to improve this paper. Warm thanks to Rachana Kametkar, Michael Gill, Jay Garfield, Mark Siderits, and Susanna Siegel for questions and discussion. An early version of this paper was presented at the 2018 NEH Institute on Self-Knowledge in Eastern and Western Philosophies, and we thank the audience for their feedback.


