Dear participants in the Cornell colloquium: my apologies for sending you such a rough draft of a paper I currently have in progress (especially, with regard to the shamefully incomplete list of references). Still, I hope to complete this within the next few months and will be very grateful for your critical feedback – Cheers, Tori McGeer

Empathy Internalized: On the scaffolding power of self-directed emotion.
WORKING DRAFT – not for citation or (general) circulation
31st October 2021

Victoria McGeer
vmcgeer@princeton.edu

ABSTRACT:

It is commonly accepted that empathetic contact with others can play a key role in supporting and enhancing our own moral agency by way of engaging our emotions in a fitting or appropriate way. I examine a particular instance of this commonsense view: that empathetic contact with others can be a powerful instigator of moral development by way of generating the self-castigating emotions of guilt, shame and remorse when we have done wrong to them. In defence of this commonsense view, I argue the self-castigating emotions can be epistemically valuable so far as they promote insight into our conduct and character that may be necessary for such development. But more problematically, these emotions can also be motivationally counterproductive for such development. To overcome this problem, I examine the conditions under which these self-castigating emotions can be managed, contained or metabolized, thereby supporting rather than defeating our self-development. My suggestion will be that forging an empathetic connection with our own erring self may play a critical role in this developmental process.

Introduction:

Discussions of empathy are generally focussed on the problem of how we engage with other minds: normally, other people’s minds; with how they think and (more importantly) feel about things; with the qualitative nature of their lived subjective experience.

For cognitive scientists and philosophers of mind, the key questions are primarily epistemic: what is distinctive about this way of knowing other minds and what are the psychological mechanisms that underpin it? To what degree are we directly in touch with what others think and feel, as opposed to generating beliefs about their thoughts and
feelings by way of simulation, imagination, projection or (unconscious) theoretically-mediated inference?

For moral psychologists and philosophers, the key questions have a different focus. If empathy provides us with a uniquely powerful way of understanding others, if it connects us in some immediate affect-involving way to how they are experiencing things in their situation, how might this figure in our moral understanding, judgement and behaviour? Does it simply provide a situationally constrained motivational boost to prosocial inclinations that are already in place (C. Batson, 1991; C. D. Batson & Shaw, 1991)? Or can it serve the deeper function of re-orienting our understanding of the moral significance of other individuals, generating or strengthening a sympathetic concern for them “for their own sake” (as Darwall (1998) would say), and thereby providing a spur to (adult) moral development?

This second possibility is the topic I mean to explore in this paper: empathy as it plays a role in significant moral development. My paper will be divided into three sections. In Section 1, I present the commonsense view that moments of empathetic contact with others can provide a spur to moral development, especially by way of generating the self-castigating emotions of guilt, shame and remorse when we do wrong to them. In Section 2, I argue that the commonsense view is on the right track insofar as these emotions can be epistemically valuable, giving us insights into our conduct and character that may be necessary for such development. However, in Section 3, I argue that the commonsense view is more problematic insofar as these emotions can be motivationally counterproductive for such development. To address this problem, I examine the conditions under which such emotions might be managed, contained or metabolized, allowing us to capitalize on the epistemic insights they provide. My final suggestion is that an agent’s capacity for empathy plays a critical role in this process, but here the empathy in question must be directed towards themselves. Hence, the title of my paper, ‘empathy internalized’. 
Section 1: Empathy for others as a spur to moral development - the commonsense view

The commonsense view that empathetic contact with others provides a spur to moral development is a well-trodden theme in the work of great novelists. I here give two examples.

Consider, first, a memorable scene from Dostoevsky’s The Brothers Karamazov. One of the novel’s more sympathetic characters, Father Zossima, is recounting the critical moment in his early life when he came to abandon his prestige-seeking career as a young army officer for the humble calling of a monk.¹ This ‘turning point’ came one evening when he struck his servant, Afanasy, twice in the face so that he was “covered in blood”. “He had not been long in my service,” Zossima continues, “and I had struck him before, but never with such ferocious cruelty. And believe me, though it’s forty years ago, I recall it now with shame and pain.” That moment is made vivid in his mind by the persisting painful image of Afanasy stoically receiving the blows, an image that likewise makes him relive the flood of shame and remorse he felt on the occasion:

*It... all was, as it were, repeated over again; he stood before me and I was beating him straight on the face and he was holding his arms stiffly down, his head erect, his eyes fixed upon me as though on parade. He staggered at every blow and did not even dare to raise his hands to protect himself. That is what a man has been brought to, and that was a man beating a fellow creature! What a crime! It was as though a sharp dagger had pierced me right through.* (The Brothers Karamazov, Bk 6, Chap. 2).

My second example of a face-to-face encounter, equally momentous, comes from the novel Middlemarch by George Eliot. It centres on the character of Fred Vincy, a rather feckless and directionless young man, who earlier in the story had convinced his hardworking friend, Mr. Garth, to guarantee a loan he was sure to pay off. But when his

¹ This scene was brought to my attention by Christopher Cordner, who discusses it in detail in “Remorse and Moral Identity” (Cordner, 2008). See too (Cordner, 2001).
schemes go awry, Fred must call on the Garths to pay his debt. Naturally, he feels terrible about imposing on them. But his encounter with the family turns out to be “more unpleasant than he had expected” when he’s brought face-to-face with how the Garths will suffer in consequence:

"I must give you the ninety-two pounds that I have put by for Alfred's premium," said Mrs. Garth, gravely and decisively... "And I have no doubt that Mary has twenty pounds saved from her salary by this time. She will advance it."  

Mrs. Garth had not again looked at Fred, and was not in the least calculating what words she should use to cut him the most effectively... But she had made Fred feel for the first time something like the tooth of remorse. Curiously enough, his pain in the affair beforehand had consisted almost entirely in the sense that he must seem dishonourable, and sink in the opinion of the Garths: he had not occupied himself with the inconvenience and possible injury that his breach might occasion them, for this exercise of the imagination on other people's needs is not common with hopeful young gentlemen. Indeed we are most of us brought up in the notion that the highest motive for not doing a wrong is something irrespective of the beings who would suffer the wrong. But at this moment he suddenly saw himself as a pitiful rascal who was robbing two women of their savings." (Eliot, 1996: Bk 2, Ch. 2)

Fred slinks away from this encounter newly awash in shame and guilt. And, as the novel unfolds, we find this moment of moral epiphany has indeed begun to work a significant change in him, enabling him (by degrees) to bring his self-indulgent tendencies under his agential control.

I use these literary examples to illustrate and highlight three key points. First, they nicely illustrate the widely held view that such moments of empathetic contact can play a

---

2 “Alfred’s premium” is the critical money Mrs. Garth’s has saved up for her youngest son’s apprenticeship. Mary is her eldest daughter, already earning money in a job she very much dislikes to help support the family.
critical role in enabling us to escape from what Eliot calls our native ‘moral stupidity’. In her view, this is a perfectly natural first-personal bias that stems from experiencing the world primarily through the lens of our own cares and concerns. In this condition, we may in fact care about other people. But their cares and concerns, and how they experience the world in consequence, remain out of focus, distorting our understanding of what it is to care for them in a way that’s appropriately tuned in to who they are as moral beings. By bringing their cares and concerns sharply into focus, empathy enables a needed re-orientation in our moral landscape.

A second key point is that, while such empathetically mediated moral epiphanies may occur under a range of different circumstances, the suffering of others is a particularly compelling stimulus for us – and even more so when we perceive ourselves to be culpably responsible for it. Our pained recognition of this fact typically and appropriately manifests itself in feelings of shame, guilt and remorse.

Finally, these examples from literature strongly imply a third critical point: that, in cases of self-perceived culpability, the self-castigating emotions (shame, guilt and remorse) are not incidental to whatever moral development and transformation these characters undergo. This is a commonly held view, not just amongst ordinary folk, but also amongst those who are professionally concerned with moral development and reform, whether as practitioners or theoreticians. And it is here I want to focus some attention.

My concern in the remainder of this paper is whether this common conception of the significance of self-castigating emotions for moral development and transformation is really on target. For there is good evidence that such emotions can get in the way of such development, as much as they may inspire it. So my question is: what makes for the difference? In the next section (2), I examine these emotions in more detail, considering a proposal, popular amongst psychologists and philosophers, that the answer to this question lies in the distinctive nature of the emotions themselves: that shame, in particular, is normatively counter-productive, and should be discouraged wherever
possible. By the end of the section, I will come to reject this proposal primarily on conceptual grounds, but with an eye to accommodating the empirical literature reviewed.

**Section 2: Self-castigating emotions and their epistemic value**

The difference between guilt and shame is very clear --- in theory. We feel guilty for what we do. We feel shame for what we are.

LEWIS B. SMEDES, *Shame and Grace*

One of the misfortunes of our time is that in getting rid of false shame we have killed off so much real shame as well.

LOUIS KRONENBERGER, *Company Manners*

I begin this section with a few words of clarification: By ‘self-castigating emotions’, I mean to designate those emotions that we typically feel, or at least feel more intensely, when we blame ourselves for some self-perceived wrongdoing. As noted in Section 1, I include on this list: guilt, shame and remorse. Some other emotions might be added to the list, such as embarrassment or regret. But these frequently lack the presumption of (culpable) wrongdoing, or at least responsibility in an extended sense, that lies at the core of guilt, shame and remorse. And it is this presumption I here want to emphasize.

---

By an ‘extended sense of responsibility’, I mean responsibility felt (and sometimes taken) for things outside of one’s agential control. As many theorists remark, shame in particular is often experienced in these circumstances. For instance, a person may feel shame for how others behave; or someone may feel shame for having been caught up in events that are morally problematic, even though their own agency is not directly implicated; or someone may feel shame for certain features of themselves over which they exercise no agential control. What seems to be common in all these experiences of shame is that people take the triggering cause to reflect on who they are in some normatively significant sense. Needless to say, many such experiences of shame are not well grounded. But others may be; in particular, the kind of shame I focus on in this paper is potentially in this category – viz., shame experienced in relation to events for which one is culpably (agentially) responsible.
Further, I call these emotions ‘self-castigating’ because I want to emphasize the way in which they are importantly connected with, though distinct from, self-blame: blame that a person directs towards themself. Surprisingly, philosophers commonly assume (generally without argument) that these emotions are simply manifestations of, and therefore equivalent to, self-blame (REFS). But I resist this conflation on both empirical and conceptual grounds. On the empirical front, studies indicate that people can experience these emotions even when they don’t judge themselves at fault, so long as they are blamed by others. This seems particularly true of shame⁴ (REFS); but it also true of guilt (Parkinson & Illingworth, 2009).⁵ This leads directly to the conceptual point that if such emotions can be felt in the absence of self-blame, they are better understood as evoked by blame, whether the blame comes from others or from oneself.⁶

Still, I label these emotions ‘self-castigating’ to stress the significance of their association with self-blame, as against merely the blame of others. For when generated by self-blame, they are likely to have a sharper and more distressing phenomenological character, as well as being more persistently robust. This should not be surprising. Once we take on the active role of blaming ourselves, we apparently accept the aptness of these self-castigating emotions, thereby finding less reason to deny or suppress them. In addition, we must bear all the discomfort of being divided against ourselves, occupying both the position of someone who actively and disparagingly judges our character and

---

⁴ By which I mean feeling ashamed versus simply feeling shamed (and therefore humiliated or degraded) by others.
⁵ This may not be true of remorse – the question has not been studied. But there is good reason for this, which I will get to below.
⁶ To elaborate a bit further on the conceptual point, I think it makes sense to distinguish between active and passive ‘reactive’ emotions – i.e. emotions, in PF Strawson’s terminology, that are particularly responsive to the attitudes and doings of responsible agents (Strawson, 1962). Strawson himself originally distinguished between self-reactive attitudes (e.g. guilt, shame, pride) as against those that are other-directed (e.g. resentment, indignation, gratitude). But, with somewhat different emphasis, he might have called these particular self-reactive attitudes, ‘passive’, since these are experienced in virtue of being the target of someone’s judgement (whether the judger is another or oneself). Equally, he might have called the (generally) other-directed attitudes, ‘active’, since these are experienced in virtue of judging someone’s actions, attitudes or character (whether the person so judged is oneself or another). (Query – better terms for this distinction? – ‘sender’ vs. ‘recipient’ RAs?)
conduct and the position of the one who is being judged. Indeed, as I will suggest in Section 3 below, this duality in our self-experience is critical to keep in mind, since it’s how we bear up under and resolve this pained division in ourselves that ultimately determines whether various self-castigating emotions are positively implicated in moral development and/or transformation.

Having made these points of clarification, I turn now to the main business of this section: an examination of the distinctive nature of these emotions themselves. How are guilt, shame and remorse to be differentiated – and what normative work might they do for us?

I begin with remorse, which generally inspires less normative controversy than either shame or guilt. Presumably, this is because remorse seems to have an unproblematically other-regarding attentional orientation. In feeling remorse, it seems I am wholly focussed on the harm or suffering I have culpably brought upon others. Of course, this involves a recognition of my own culpability, but my concern is not with myself -- or even my past behaviour. It may be with my current or future behaviour, but only as that behaviour can alleviate or atone for the harm I have caused. Hence, the normative value of remorse seems unquestionable. Unsurprisingly, then, it is often

---

7 cf. Adam Smith – “when I endeavour to examine my own conduct, when I endeavour to pass sentence upon it, and either to approve or condemn it, it is evident that, in all such cases, I divide myself, as it were, into two persons; and that I, the examiner and judge, represent a different character from the other I, the person whose conduct is examined into and judged of. The first is the spectator, whose sentiments with regard to my own conduct I endeavour to enter into, by placing myself in his situation, and by considering how it would appear to me, when see from that particular point of view. The second is the agent, the person whom I properly call myself, and of whose conduct, under the character of a spectator, I was endeavouring to form some opinion. The first is the judge; the second the person judged of. But that the judge should, in every respect, be the same as the person judged of, is as impossible, as that the cause should, in every respect, be the same with the effect.” (Smith, 1759/1982: Chapter I, Part III.)

8 This is not an easy topic: The boundaries of these emotions are not well established in folk-psychology; and trying to bring some regimentation to these concepts is bound to be controversial. Still, we can make no headway on the normative questions without taking some stand on the differentiating features of these emotions; so this is what I intend to do, partly by means of philosophical reflection, and partly through leaning on empirical studies (especially with regard to shame and guilt; remorse has not been subject to the same degree of empirical attention, perhaps because it inspires less normative controversy).
singled out, in both practical and theoretical contexts, as the emotion we hope offenders will (come to) feel in the aftermath of wrongdoing.  

But now what of shame and guilt? While both these emotions are generally more self-oriented than remorse, there are certain key differences between them that have led a number of philosophers and psychologists to question the normative value of shame in particular. Amongst these key differences (summarized in FIGURE 1), the following are generally acknowledged: (a) shame involves a negative evaluation of the person’s self as a whole, whereas guilt involves the negative evaluation of their acts; (b) shame is linked more closely to a violation of a person’s values or ideals, whereas guilt is linked to the violation of prohibitions (Flanigan REF, but see too Gibbons 1990, Gilbert 1997, Leary 2000 for a different view – cited by Harris 2003); (c) shame involves feelings of threatened self-identity, whereas guilt involves no such preoccupation; and, finally, (d) shame involves more focussed attention on the self, as well as heightened vulnerability to condemnation from others. REFS- very long list!, esp including Tangney et al

More controversially, shame and guilt are purported to have different action-tendencies, as well as downstream affective and behavioural consequences: In particular, shame seems to motivate more anti-social self-defensive behaviours, such as avoidance, withdrawal, denial, deflecting blame on to others and aggression. By contrast, guilt seems to motivate more pro-social self-deprecating behaviours such as apology, admitting wrongdoing, making amends, and committing to do better in future. In keeping with this, shame is not only self-focussed in an attentional sense; it purportedly generates a more obsessive esteem-based concern with self-image and reputation. By contrast, though guilt is attentionally focussed on the moral probity of one’s own deeds, it purportedly generates a more empathetically mediated, other directed concern -- i.e. remorse. (REFS – both pro and con).

---

9 Interestingly, remorse is generally selected by moral philosophers as the emotion wrongdoers ought to feel – and that we try to make them feel by way of our blame. See, for instance, Fricker for a particularly clear statement of this view (Fricker, 2016). Other philosophers that put a high premium on remorse: Duff, Bennett, Gaita, Cordner.... Shame/ guilt are rather skipped over, though possibly viewed as necessary way-stations and/or accompaniments...

10 For a normative attack on guilt, see G. Harman (REF)
Unsurprisingly, it is these findings – particularly, those I have labelled ‘more controversial (highlighted red in the summary table below) -- that have led many to embrace the conclusion that shame is a normatively suspect emotion, one that positively inhibits moral behaviour and moral development. Guilt, by contrast, is something we can celebrate – and presumably encourage wrongdoers to feel to the exclusion of shame.

<table>
<thead>
<tr>
<th>Shame</th>
<th>Guilt</th>
</tr>
</thead>
<tbody>
<tr>
<td>• involves negative evaluation of one's self</td>
<td>• involves negative evaluation of one's acts</td>
</tr>
<tr>
<td>• assoc. w/ violation of one's values/ ideals</td>
<td>• assoc. w/ violation of (moral) prohibitions</td>
</tr>
<tr>
<td>• assoc. w/ feelings of threatened self-identity</td>
<td>• no feelings of threatened self-identity</td>
</tr>
<tr>
<td>• heightened vulnerability to others’ condemnation</td>
<td>• no overweening concern with others’ condemnation</td>
</tr>
</tbody>
</table>

• purportedly motivates more anti-social self-defensive behaviours \(\rightarrow\) avoidance, withdrawal, denial, deflecting blame on to others and aggression;

• purportedly generates more self-directed (esteem-based) concern with self-image and reputation.

• purportedly motivates more pro-social self-deprecating behaviours \(\rightarrow\) apology, admitting wrongdoing, making amends, and committing to do better in future).

• purportedly generates more other-directed concern \(\rightarrow\) i.e. remorse

Figure 1

As just noted, this normative recommendation is fairly widespread: that we should aim to expunge shame from our moral-psychological repertoire in favour of guilt. But I am suspicious of this normative recommendation even granted the constellation of unhappy features various researchers have found to be associated with shame. I have two reasons for this: one practical and the other normative.

The **practical reason** can be stated succinctly: We may not be happy with the structure of human emotions, but they are what they are – regulable to some extent, but not (I suspect) modifiable in the way envisioned here. Shame is a pervasive feature of our moral psychology, likely to be triggered under a variety of conditions, both apt and inapt – and whether we actively blame wrongdoers or not (cf. the cases of Fred and
Zosimma above).\textsuperscript{11} So wishing it away is not a particularly helpful normative stance to take -- particularly in cases of serious moral wrongdoing, where research suggests that offenders’ experience of shame and guilt are deeply intertwined, apparently occurring “as a single response” (Harris, 2003). So I suggest a more constructive normative stance is one of figuring out how shame experiences might be managed more effectively, so that shame can join with guilt in promoting rather than inhibiting moral development/ transformation.\textsuperscript{12}

The deeper normative reason I have for resisting the idea of expunging shame from our moral psychological repertoire builds on the characterization of shame given above. For it does suggest that shame can do a kind of normative work that guilt simply doesn’t do – and this notwithstanding the liabilities often associated with it that psychologists rightly highlight. Indeed, in my view these liabilities underscore the normative power of shame – an emotion that makes us turn against ourselves, undoing the satisfaction, or perhaps simply comfort, that we feel in our own self-identity. So it’s no wonder that feelings of shame should evoke a number of defence mechanisms. And yet this kind of self-dislocation is sometimes precisely what’s needed to initiate a process of genuine moral development.

Consider, again, the case of Fred Vincy. As Eliot makes clear, Fred certainly feels a burden of guilt prior to visiting the Garths. Indeed, that’s what prompts his visit: his guilty recognition that his ill-judged behaviour has put them in a bad situation, and his guilty understanding that he owes them an apology, as well as assurances of his intention to make good the loss he has brought down upon them. So he feels that he’s doing just what an ‘honourable young man’ should do in the situation – viz. face up and acknowledge responsibility for the harm he has caused. Consequently, en route to

\textsuperscript{11} Notice that in the two literary examples I gave, shame was endogenously manufactured – i.e. not as the result of anyone else’s blame, suggesting that there are limits to how much our blaming practices could alter this basic feature of our moral psychology.

\textsuperscript{12} Here I am presuming that guilt really is as normatively unproblematic as researchers suggest. My own view is that guilt experiences might need some management as well in order to promote rather than inhibit moral development. For discussion in relation to the problem of addiction, see (Snoek, McGeer, Brandenburg, & Kennett, 2021).
making them his apologies, he does not feel *ashamed* of himself, of the kind of person he is.

But all this changes when Fred is brought face-to-face with how his feckless self-interested behaviour truly impacts upon the Garths: “... *he suddenly saw himself as a pitiful rascal who was robbing two women of their savings.*” Admittedly, this moral epiphany involves reconstruing the nature of what he’s done, revising and presumably deepening his feelings of guilt in consequence. But this reconstrual is itself brought about by a more fundamental change in Fred’s vision of himself: no longer the morally honourable young man who has unintentionally burdened his friends, but rather a morally shiftless young rogue too caught up in pursuing his own interests to care about the damage he might do to others. Honourable young men do not rob two women of their savings; shiftless young rogues do. Fred now sees himself as this ‘pitiful rascal’ and is covered in shame. His apologies, moreover, bring him no relief. Instead, burdened by this newfound vision of himself, he soon hurries away “feeling confusedly that his being sorry was not of much use to the Garths” (Eliot, 1996: Bk 2, ch. 2).

Why think the kind of shame that Eliot here portrays is morally valuable? It certainly does not empower Fred to make better or deeper apologies to the Garths, or suggest more significant ways of making amends – at least right away. On the contrary: he is motivated to ‘hurry away’, preoccupied by his own inadequacy and the pointlessness of any apologies he might make to them. And yet we readers are now beginning to feel some hope for Fred. In this moment of shame, he finally demonstrates some genuine insight into the nature of his own moral character. And he is not just giving himself a pass. Instead, he is allowing that insight to reverberate through the fabric of his own self-conception, replacing a comfortable vision of himself and his activities, with a much bleaker vision of who he truly might be: the *kind* of person who could do *that* to his friends. And while this process certainly involves a heightened degree of self-focus, it’s surely the kind of self-focus—viz. pained attention to one’s own
morally problematic traits and dispositions – that’s pre-requisite to any genuine moral development.\textsuperscript{13}

Let me close this section, then, by suggesting that, however problematic shame experiences can be, we go too far in simply dismissing shame as a normatively corrosive emotion – to be suppressed or discouraged at all costs. More strongly, far from there being good or bad self-castigating emotions, I suggest that all of them have an important epistemic role to play in keeping us morally on track. For, as noted already, these emotions are distinguishable in terms of how they focus our attention: In feeling remorse, we focus on the harm we’ve done to others; in feeling guilt, we focus on the transgressive nature of our deeds; and in feeling shame, we focus on aspects of our own character (our traits and dispositions) that support such unhappy behaviour. But this is just to say that all three emotions focus our attention on features of our actions that are relevant to improving our attitudes and behaviour going forward.

Indeed, I am tempted to make a stronger point still: that, in certain cases, it’s critical that all three emotions work in combination, modulating how each are experienced. For this is what enables us to truly grasp the moral significance of what we do. Consider, again, the case of Fred Vincy: it’s only when he experiences the ‘prick of genuine remorse’ that his guilt and shame take on a proper depth and focus, giving him new insight into the nature of what he has done, as well as into his own characterological weaknesses. And it’s only when he experiences shame and guilt in the right manner and measure, that his remorseful attention to the suffering of others takes on a proper depth and focus as well, staying with him and supporting his stumbling steps towards moral development and reform.

\textsuperscript{13} Of course, I am here presuming that people do have stable traits and dispositions – stable enough to make this kind of recognition apposite. For challenges to this view, see (Doris, 2002). (\textit{Query: to what extent do I need to defend this presupposition for the purposes of this paper?})
Section 3: The scaffolding power of self-castigating emotions - narrative reconstruction & self-directed empathy

Take away my past! I know you can’t take away the past, but I need to learn how to cope with it. (41-year-old, eighth prison sentence)

S. MARUNA & D. RAMSDEN, “Living to tell the tale”

As I said in Section I, we commonly assume, as a matter of folk-psychology that the self-castigating emotions of shame, guilt and remorse play a critical role in moral transformation and development. So far I’ve defended the thought that they are epistemically valuable so far as they bring home to us the moral significance of features of others, of our actions, and of ourselves that we had erstwhile discounted or ignored. But are they really as valuable motivationally, as they are epistemically? Can they play a positive scaffolding role in an agent’s psychic economy, prodding her to do better in her life going forward? In this section, I want to address this question by focussing on the problem of offender reform in a criminal context.

I do so for three reasons. First, even though literary examples are helpful in illustrating various themes, the problem of desistance from crime (or ‘going straight’) provides unvarnished insight into the moral-psychological reality of what it takes to undergo genuine and substantial moral transformation. Secondly, the problem of desistance has been extensively studied, so there is rich material of both an empirical and conceptual nature upon which to draw (so-called ‘desistance literature’). But my third and most important reason is this: it seems to be widely assumed, as a matter of criminal justice practice, that self-castigating emotions do have positive motivational effects. Hence, the enthusiasm for so-called ‘shaming punishments’ in certain criminal jurisdictions (although I agree with Jeffrie Murphy that such enthusiasm is quite likely to be fed by “a kind of smug and mean-spirited vengeance with tendencies to lapse into
arbitrary cruelty” (Murphy, 1999, p. 338). Still, more humane practices or therapeutic interventions might aim to “responsibilize” criminal offenders (Garland, 1997) by inducing self-castigating emotions, on the assumption that such emotions will be valuable motivationally as well as epistemically.

But is this assumption well grounded? I think the evidence here is mixed – and the reason is clear. It is one thing for individuals to ‘experience the sting’ of self-castigating shame, guilt and remorse; it’s quite another for this pained acknowledgement of one’s own culpability to drive the more demanding process of self-transformation – particularly in cases of serious moral wrong, where these feelings are bound to be particularly deep and intense. Indeed, experiences of ‘atonement’ (as we might call it) seem to come in two flavours. The first involves continued and continuing self-flagellation; the second involves a more constructive focus on making amends for the harm one has done. The first involves an inwardly directed preoccupation with one’s bad (and seemingly irredeemable) self; the second, an outwardly directed concern with doing and making better.

If shame, guilt and remorse are implicated in both ‘atonement’ trajectories, why should there be such differences between them? While some may yet insist that the difference surely lies in the nature of the self-castigating emotions themselves, I suggest it lies elsewhere – in the character of an individual’s self-blame. For, as I’ve argued elsewhere (McGeer, 2013, 2018; McGeer & Funk, 2017; Snoek et al., 2021), blaming

---

14 Importantly, as he observes, such shaming punishments “… do not engage and rebuild the core of the moral self but simply add extra punitive burdens and inconveniences (some of them quiet grotesque) to the criminal’s post-conviction life” (Ibid, p. 338)

15 This seems to be Sartre’s view, endorsed by Dan Zahavi, who sees the following fundamental problem with shame (versus guilt): “The future is lost, and the subject is fixed in the present moment. As Sartre writes, in shame I experience myself as trapped in facticity, as being irredeemably what I am (rather than a someone with future possibilities, as someone who can become otherwise), as defencelessly illuminated by an absolute light (with no protective privacy). Whereas guilt is primarily focussed on the negative effects on others, includes a wish to undo the deed, and might motivate reparative actions, the acute feeling of interpersonal shame does not leave room for the exploration of future possibilities of redemption” (Zahavi, 2014, p. 223). As will be evident here, I agree that ‘fixity’ in the subject’s experience of themselves is the critical problem; but I think it is the character of their self-blame that lies at the root of this.
itself is an activity that comes in two flavours; and that will be true, whether it’s directed towards others or directed towards ourselves: Retributive blame focuses on unilaterally condemning wrongdoers (as in, ‘an eye for an eye’); Scaffolding blame, by contrast, focuses on calling wrongdoers to account, insisting on their undertaking some serious self-examination in order to understand and amend their behaviour going forward. Analogously, I suggest, self-blame can be undertaken in a retributive condemnatory mode – or in the scaffolding mode of calling oneself to account. And when it occurs in the retributive mode, a number of evils arise.

Why should this be the case? The answer lies in the structure of retributive blame that is both non-dialogical and focussed on the past. Its aim is simply to deliver an appropriately condemnatory response to offenders that is merited by their bad deeds and (ultimately) bad character. Once this deserved response is delivered -- and some would say, understood -- the goal of retributive blame is met (Gollwitzer & Denzler, 2009). In this sense, it is unidirectional and so essentially non-dialogical. Offenders are simply expected to accept this condemnatory response as their due – i.e. their ‘just deserts’.

Consider now what happens when self-blamers engage in this condemnatory activity: an activity they regard as merited by what they have done. I suggest they are emotionally primed to accept the validity of their own retributive blame by embracing -- indeed, wallowing in -- their bad character. Fred Vincy is a perfect exemplar of this tendency, at least initially. “I am a blackguard,” he says to Mary Garth in the first throes of his shame, guilt and remorse: “You can never forgive me”. Indeed, as a self-professed blackguard, Fred now sees himself as doomed to live out his life in accordance with (Maruna calls) a “condemnation script” (Maruna, 2001) -- fated to be good for no one, not even himself. He is powerless, so he feels, to escape his own nature. His traits and dispositions are a fixed part of himself – traits and dispositions that he can neither avow from a first-person point of view, nor disavow. As he sees it, the only thing left is simply to own or acquiesce in his tarnished self, submissively accepting – and indeed seconding -- the condemnation that rightly comes his way.
Unsurprisingly, when we look at the actual desistance literature, this self-punishing stance is generally associated with an obsessional self-focus, low self-esteem, depression, lassitude, self-pity, emptiness and/or existential despair (Pickard, Maruna). More deeply, it constitutes an abrogation of the individual’s first-personal agential capacity to take responsibility for his/her own traits and dispositions and make them otherwise (McGeer, 1996, 2008; Moran, 2001). It is, in effect, an indirect way of shirking responsibility for the tarnished self a person supposedly accepts as their own (Pickard).

The alternative, more positive ‘atonement’ trajectory involves what I have called self-blame in the ‘scaffolding mode’. Unlike retributive blame, the essential structure of scaffolding blame is dialogical and future-focussed, presupposing a capacity in the self being blamed to respond proactively, changing and/or developing under the spur of being called to account. An individual who blames herself in the scaffolding mode is already primed to see herself, not as a character fixed in stone, but as an individual whose traits and dispositions are hers to modify and manage. Guilt, shame and remorse focus her pained attention on where work needs to be done. But these emotions also serve as prods to undertake a process -- perhaps a long and difficult process -- of self-development, the undertaking of which itself naturally generates more positive self-directed emotions as the process unfolds (emotions such as hope-for-oneself, self-trust, and self-forgiveness).

Distinguishing between these two different modes of self-blame is all very well in theory; but in practice they are often intertwined. Indeed, in cases of serious wrongdoing, retributive self-blame is likely to dominate, especially in the face of institutionally backed external condemnation. Thus, criminal offenders are particularly prone to embrace condemnations scripts, seeing their lives as simply “wasted” and beyond redemption (Maruna & Ramsden, 2004). What, then, might allow such agents to move into a different register of self-blame, whereby they come to operate as reform-minded self-scaffolders, rather than simply acquiescent self-punishers?
In the literature on rehabilitation and recovery, narrative therapy is lauded as particularly useful in this regard. As one practitioner, E.B. O'Reilly, explains:

“Narrative is not a cure, but it is a method, a path towards redemption. Redemption lies in … a better understanding -- an improved epistemology -- including the development of a talent for recognizing counterfeit, seeing through duplicity, and resisting snares and seductions” (O'Reilly, 1997, pp., p. 65).

This ‘improved epistemology’ comes about by way of encouraging persistent offenders to embrace a new conception of who they are -- a new ‘self-identity’ -- in the context of developing a more ‘generative story’ of their lives. The successful ‘desistance narrative’ is one that enables these individuals to cast-off, or disown, their former deviant self -- the self that engaged in criminal (or self-destructive) behaviour -- and embrace a new way of life by living up to their newly constructed self-image. As another advocate of narrative therapy, Barry Vaughan, writes:

“the self … is now being defined in terms of attestation, being constant to some future ideal self that acknowledges yet disavows past actions” (Vaughan, 2006, pp., p. 396)

Critically, this narrative process is one that explicitly acknowledges and celebrates the agent’s new found first-personal capacity for “self-constitution” (Ricouer, Korsgaard, Moran, McGeer). That is to say, the new self is not only narratively constructed; it is constructed as a self newly empowered to take responsibility for self-making -- i.e., for shaping its own projects, as well as its own traits and dispositions. The past self is thus disowned, not simply for actions once performed and now disdained, but also for its willingness to acquiesce in habits of self-excusing passivity -- the old ‘epistemology’ that is also now disdained. So creating a new ‘self’ and discarding the old seems to be a critical trope in allowing individuals to move forward with their lives; to forego punishing self-blame in favour of a more constructive and self-empowering process of calling themselves to account.
But here we might ask, just how radical does the break between these old and new ‘selves’ have to be in order to support a stable process of self-transformation and reform? For as Maruna and Ramsden point out, one of the challenges of genuine desistance is to develop a sufficiently stable new conception of the self so as to withstand “predictable setbacks, obstacles and disappointments in the life of a recovering person” (Maruna & Ramsden, 2004, pp., p. 138). And here it may seem that the more ‘distance’ individuals are able to put between their old and new selves the better.

In her work on personal identity, Marya Schechtman (2001) claims that individuals achieve a genuine break with their past selves (and hence a new identity) when they no longer have ‘empathetic access’ to the way they once thought and felt (p. 102). As she puts it, such individuals may have a ‘cognitive recollection’ of being the person who once engaged in a certain way of life; but they are ‘profoundly alienated’ from that way of life so far as they “… cannot recapture the passions, emotions, likes and dislikes…” of the self they once were (p.101). Past experiences are thus effectively ‘knifed off’ from the present; and individuals may speak of themselves as effectively ‘born again’ (Maruna & Ramsden, 2004, pp., pp. 138-139).

This suggests one straightforward strategy offenders might adopt in the service of achieving genuine reform: the ‘born again’ strategy. In keeping with Schechtman’s characterization, it allows them to acknowledge their past self, biographically speaking, while refusing to engage empathetically with how such a person might have thought and felt. That self is simply condemned and abandoned. Instead, they fully embrace a new way of being by projecting themselves entirely into the future, disowning any lived experiential connection to the person they once were.

Attractive as this may sound, the literature on rehabilitation and reform suggests, perhaps counter-intuitively, that this is not the optimal path to stable self-transformation and moral reform (Maruna, 2001; Maruna & Ramsden, 2004; O'Reilly, 1997; Rotenberg, 1987). On the contrary, the narratives of successful ex-offenders and recovering addicts
display a “strong conserving trend” in how these individuals conceptualize their new identity (O'Reilly, 1997, p. 152). Importantly, their past self is not one that is straightforwardly disowned; it is rather one that is empathetically re-visioned and reclaimed by way of reformed individuals proactively asserting that the self they ‘truly’ are -- and the self they have always been -- is more than the sum of the dispositions and traits that led to past dysfunctional behaviour. Let us call this the ‘once and future self’ strategy individuals may adopt in the service of achieving genuine reform.

How precisely does this strategy work – in particular, how is this ‘conserving trend’ commensurate with substantial change and development? It’s important to stress that there is a newly emerging self, one that can ‘take’ responsibility for traits and dispositions that are now repudiated. But, critically for these individuals, they see this newly empowered self as one that draws upon features they always had – i.e. features that were present, though admittedly not dominant, in past ways of being. Thus, successful desistance narratives seemingly depend upon individuals ‘re-biographing’ or ‘correcting’ their past, “for the sake of psychological continuity and cognitive congruity” (Rotenberg 1987, p. 49). In such narratives, reformed individuals certainly acknowledge their past dysfunctional activities and supporting dispositions; but they are eclipsed by memories of a once and future ‘better’ self now stepping onto centre stage. As forensic psychologist, Mordechai Rotenberg puts it:

The failing parts in a person’s history are contracted while the reinterpreted parts are expanded to create a more congruent life story dialogue between the future-oriented present new ‘I’ and the past ‘thou’. (Rotenberg -- p. 65).

Having explained the mechanics of the ‘once and future self’ strategy, we should now consider why it is practically efficacious, especially in comparison with the more straightforward – indeed, epistemically less suspect -- ‘born again’ strategy. What is the value in placing so much emphasis on re-making or ‘correcting’ (some might say

16 Footnote to Jones on remaking the past…(Jones, 2008)
‘distorting’) the past? How can projecting a new-found conception of the self back in time aid in genuine moral development/ transformation?

I conclude with two speculative observations. First observation: it seems that, even when individuals want or need to leave part of themselves behind, the integrative demands of coherent, well-functioning agency imposes on them a powerful countervailing need to maintain empathetic contact with who they once were: clearly, not as individuals whose dysfunctional values and lived experience they can still own in a first-personal sense (i.e. avow); but at least as individuals they can still make sense of (i.e. access or understand) in an immediate visceral way, even from their current disavowing perspective.17 As one ex-offender put it: “I know you can’t take away the past, but I need to learn to cope with it”. And the coping mechanism here -- the truly generative coping mechanism -- is one that refuses to make a stranger of the past self, but instead makes peace with it through empathetic connection and understanding.

Second observation: This coping mechanism brings with it a significant additional advantage. For in addition to bolstering their sense of coherent, integrated agency through time, maintaining empathetic contact with their past selves allows reformed (or reforming) agents to maintain better managerial control of the traits and dispositions that once led them astray. For such control depends on retaining a vivid understanding of what makes the old way of life so easy to give into -- the ‘snares and seductions’ of the bad epistemology these agents now forswear. Reformed (or reforming) individuals may thereby retain a sense of vulnerability to morally dysfunctional ways of thinking and being. But, ironically, it’s precisely this on-going sense of (empathetically maintained) psychological vulnerability that gives them greater resilience in adhering to their reformed way of life, as they face the setbacks and obstacles that come their way.18

17 So I disagree with Schechtman’s conception that empathetic access requires endorsement/ approval of past emotions, experiences, ways of thinking… REF here to Kennett also on the problems of cognitive dissonance.
18 Thanks to Hanna Pickard for discussion of this point.
Thus, I reach the seemingly paradoxical conclusion of this paper: securing the
good of significant identity-changing self-development is best accomplished by
maintaining an empathetic bridge between past and present – a bridge that enables
individuals to ‘own’ the very past self they are now committed to disowning.
References: (incomplete)


