

This is an uncorrected pre-proof, to appear in *Social and Legal Studies*, citations should be to the final published version please.

What Punishment Expresses

Craig Reeves

It is much easier to show that punishment has a symbolic significance than to say exactly what it is that punishment expresses...

(Feinberg, 1965: 402)

1. Expressivism – Philosophical and Psychosocial

The view that punishment should be understood as an essentially expressive practice and evaluated as such has been influential in both the normative philosophy of punishment and in psychosocial penal theory,¹ though about this view, as about many issues, there has been relatively little communication between those two traditions. Philosophical and psychosocial expressivist accounts of punishment have developed more or less independently, as if they had little to do with one another. This is a pity because expressivism is a promising conceptual account of punishment that could, if cashed out systematically in line with a realistic moral psychology, allow us to deepen our critical understanding of punishment in liberal modernity.

The crucial question for expressivist accounts is *what* it is that punishment expresses, and philosophical and psychosocial accounts tend to offer very different answers to this question. Philosophical accounts typically tell a moralised story: punishment expresses deserved or warranted (in the good case, at least) 'moral condemnation' or 'censure', and the moral emotions – resentment, disapprobation, outrage – that properly accompany it (Feinberg, 1965; Murphy and Hampton, 1988; Duff, 1986 Hampton 1992). To psychosocial theory, this will seem superficial. It focuses on what our practices of punishment *say* they express, what they avow, and on their manifest rather than their latent expressive content (See Freud, 1953). And this, psychosocial theory may say, is to operate with a naïve conception of expression which conflates symbolic expression with mere signification. One consequence of this is that philosophical accounts can struggle to say anything about the expressive significance of concrete punishment itself – that is, of i) hard treatment, and of ii) its various concrete socio-historical forms. These are typically understood as simply arbitrary conventional signs of censure – merely 'our particular symbols of infamy,' (Feinberg, 1965: 421) conceived voluntarist or decisionistic way: we have simply chosen

¹ In philosophy, Feinberg and Hampton are the most influential exponents, and arguably the view has its roots in Hegel (1821/1942); see Farnham, 2008.

these conventional 'symbols' and could choose others (see Hampton, 1992, and a critique of Hampton's position in Gert et al, 2004). Hence conventionalist expressivism shines little light on the centrality of hard treatment in penalty generally, let alone on the specific expressive meanings of the 'painful symbolic machinery' (Feinberg, 1965: 420) that we happen to have in fact developed in modern penalty.

Now this naïve conception of expression must, psychosocial theory may say, presuppose an unrealistic moral psychology – a moralizing psychology, which simply builds the psychology out of the moral categories rather than offering a genuinely psychological account of them. Psychosocial accounts offer deep interpretation of the latent content expressed in punishment that do not simply accept uncritically the *manifest* moral story of retributive penalty, that punishment expresses deserved condemnation in response to wrongdoing, but try to be sensitive to the underlying dynamics at work. They find such items as solidarity (Durkheim, 1893/1997), aggression, fear, and guilt, and their distortion, rationalisation or disavowal (Nietzsche, 1887/1998; Mead, 1918/1964) as the real, deeper, *latent* expressive content of punishment.

Recent psychoanalytic psychosocial work in this area has been promising (Garland, 1990; Maruna et al, 2004; Gadd 2007) but as yet remains rather less than fully developed and systematic. Selective in their interpretive use of disparate psychoanalytic concepts of unconscious drives, emotions and defences, they feel *ad hoc* and eclectic. But we should want to be able to make sense both of the *unities* that are found in constellations of expressed materials, and of the *antinomies* that exist between different such constellations. We should want to know why and how the constellations of specific emotional and defensive elements in punishment hang together in the ways they do, and why others cannot co-exist stably, always leading to dissonance and oscillation.

More fundamentally, such accounts may, to philosophical expressivism, seem suspiciously like a 'hermeneutics of suspicion' that, inspired perhaps by a Nietzschean take on Freud, is bound to entail reductively explaining away all ethical values and normative phenomena. If we can boil down punishment's manifest moral expressions to deeper, seedier latent ones, that – so the objection goes – must mean we can do the same for all moral phenomena, so that 'when, what we fondly believe to be reasons are unmasked, all that remains is blind causal processes,' (Brandt, ms.: 3) and then there is no normative purchase left on the world, and hence no grounds of critique either. Then it may seem such approaches are merely selectively debunking and pathologizing what they don't like (see Tetlock, 1994). That can lead Maruna et al., for example, to reason that if punishment or punitiveness is to be psychoanalytically interpreted, 'to be perfectly balanced' non- or anti-punitiveness should also be psychoanalytically interpreted in terms of 'unconscious reasons', and seemingly to infer from this that 'sympathy and concern for offenders is no more right or wrong than punitiveness.' (2004: 282-3) Through a Nietzschean lens, psychoanalytic psychosocial theory may appear to lead to the normative dead ends of such puzzling statements.

Hence, while philosophical expressivism might seem psychologically superficial and naive,

psychosocial expressivism might seem normatively muddled. So, though expressivism is promising conceptually, philosophical accounts of what punishment expresses tend to idealise it because they presuppose a naïve view of expression and an unrealistic moral psychology, while psychosocial accounts engage in less naïve, deep interpretation of punishment's symbolic expressive content, but may seem to presuppose a reductive moral psychology that leads to global scepticism. Hence not only do they seem in conflict, but moreover neither seems a promising basis for a critical understanding of punishment. It might seem that they are doomed to talk past one another, and that expressivism is of no interest for a critical theory of punishment.

But we need not draw such a pessimistic conclusion. The core analytical thesis of philosophical expressivism – that punishment is essentially an expressive practice and should therefore be comprehended and evaluated in light of that – is eminently plausible, but it needs filling in with a psychologically plausible interpretive account of *what* punishment expresses. And to begin with such an account we will need above all a realistic moral psychology that avoids both naïvety and scepticism: that is, a genuine *moral* psychology, not a *moralizing* one, yet nor a *reductive* one (Williams 1995: 202-205; Lear, 2017: 33-4). If done properly, such an expressivism could offer us a cogent understanding of punishment as an expressive practice which is capable of grounding genuine normative criticism that is both potentially radical and ethically serious.

Now psychoanalysis can be taken in a reductivist direction or in a moralizing direction, but need not be taken in either. Indeed, the British psychoanalytic tradition seems at its best to avoid both pitfalls, offering just the sort of *moral but neither moralizing nor reductive psychology* we need if we are to develop a more systematic understanding of what punishment expresses. I want to outline the claims of one of the major thinkers of the British tradition, Melanie Klein,² and then to sketch, in the broadest strokes, how an account of what punishment expresses informed by a Kleinian moral psychology might look, and to explain why such an account might be thought to provide just the sort of psychosocial filling-in of expressivism that we need if expressivism is to advance our understanding of punishment in a way that is critical and true – which must amount to the same.

2. The Priority of Unconscious Phantasy

Claims that punishment expresses helpless anxiety (Robinson and Gadd, 2016), aggressive impulses of sadism or self-assertion (Mead, 1918/1964; Carvalho and Chamberlen, 2017), envy and masochistic guilty desires for punishment (Garland, 1990; Maruna et al, 2004) or guilty feelings of complicity with the social causes of crime (Garland, 2001; Maruna et al 2004), along

2 For reasons of manageability, I restrict myself to Klein's middle period ideas, that is, before around 1957 when the concepts of envy and gratitude become central (Klein 1957/1975). The implications of the latter development I leave for another time.

with their rationalisation, disowning or disavowal, are suggestive, but offered up unsystematically they give us little in the way of a deeper grasp of the symbolic unities and antinomies of penalty. This is because they deploy abstractions from a psychoanalytic conception of soul (*psyche*), the core of which is, as Lear has argued, not the idea of an irrational unconscious mind, but rather a distinctive conception of unconscious mindedness. Freud's account of unconscious mind is not a theory of *an* unconscious mind, lurking beneath the conscious one, but a theory of a different and distinctive *form or 'principle of mental functioning'* in the mind that is contrasted with, underpins and interacts with conscious, rational thought: *unconscious phantasy* (Lear, 1998; 2017). Phantasy is a form of mental activity characterised by its own (in a sense) 'logic', a kind of order, not 'logical' in the sense in which conscious thought is logical, but not simply 'irrational', 'instinctual', mere brute nature either. Rather, it has a rationality, of a sort, of its own: it is imagistic, experiential, (proto-)narrativist, associative, metaphorical, emotional and bodily, rather than strictly conceptual, propositional, inferential and abstract. Psychoanalysis can be seen as filling out and extending Aristotle's conception of the 'non-rational nature in the soul' distinct from rational thought but operating according to its own distinctive, strange yet meaningful principle of activity, and relating to conscious rational thinking in complex entangled ways (Lear, 2017: 35).

Melanie Klein takes up Freud's thought and extends it.³ Much as cognitive functioning develops from simpler to greater complexity, phantasy has its own developmental principle, in the sense that from simple phantasy operations more complex ones become possible,⁴ and this development takes place through the complex interplay of particular narrative phantasies and the building up through them of an *internal world* that is built up from early and later phantasy activity and comes in turn to frame and constrain it but also to create qualitatively new possibilities. This developmental principle is an essential tendency to develop and grow that can of course go wrong – become stalled, inhibited, derailed – in various ways. Phantasy is then a distinctive *form of mental functioning*, with its own distinctive developmental phases. Klein calls these phases 'positions' and there are two: a primitive, 'persecutory' form, and a more developed, complex 'depressive' form. They involve fundamentally different organisations of an inner world of objects and object-relations, and of the 'anxieties and their vicissitudes' (Klein, 1946/1975: 3). Development from the one to the other is a psychological achievement that, as with the growth and maturation of an organism, involves the development of a more highly organised and differentiated degree of complexity and the acquisition through more developed organisation of more advanced powers.

3 For general introduction, see Segal (1979); Mitchell and Black (2016: ch. 4); Spillius (2011). For philosophical reconstruction and defence see Wollheim (1984) and Gardner (1993).

4 In this sense, we might see Klein as doing for unconscious mindedness what Piaget and Kohlberg attempt to do for conscious cognition, though there is a crucial difference: Kohlberg is vulnerable to the objection that he has built specific, controversial ethical assumptions into his developmental account, i.e. that his is a moralizing psychology (see e.g. Gilligan, 1982), whereas this is arguably precisely what Klein's position, at least in principle, avoids doing.

3. The Fragmented World of the Persecutory

The developmental story goes something like this. 'From the beginning object-relations are moulded by an interaction between introjection and projection,' the most rudimentary phantasy operations, and 'between internal and external objects and situations.' (Klein, 1946/1975: 2) As a form of thinking that is not separated off from bodily experience, the infant's experience of bodily interactions are manifest in phantasy: introjection is the mental, phantasy, interpretation of taking things into the body – the archetype being of course the introjection of the satisfying 'good breast': the infant phantasises taking inside itself the good, satisfying part of the parent (it could be a breast or a bottle), a process that manifests the wish that the 'the loved object may be preserved in safety inside oneself' (Klein, 1935/1975: 264). Through such introjections the infant begins to build up an *unconscious inner world* of '*internal objects*' felt to persist inside the infant and to retain their essential characteristics. But the infant will inevitably encounter frustration, and thus aggressive rage at the frustrating parent, and this rage is 'projected onto' the world. For at this stage there is no firm dividing line between what's in the infant's mind and what's going in the world: the infant hasn't developed the capacity to 'contain' its emotions (Bion, 1967: 106) so that they can appear 'as much external as a clap of thunder.' (Winnicott, 1960/1965: 141) The world is now coloured by the infant's rage, 'and so it comes about that it develops a phantastic and distorted picture of the people around it,' experiencing them as frightening, rageful, persecuting figures. The infant's response to these persecuting figures is phantasies of attacking and destroying them by sucking, biting chewing and eating them up. The corollary of these phantasies is that the bad persecutors are now felt to have been introjected and 'disintegrated' into a host of 'dangerous fragments', so that the internal world consists increasingly of 'multitude of internal persecutors' (Klein, 1935/1975: 272). The spiralling of projection and introjection organised by splitting continues to reinforce this nightmarish inner world: 'the projection of a predominantly hostile inner world which is ruled by persecutory fears leads to the introjection – in 'taking-back' – of a hostile external world; and vice versa.' (Klein, 1946/1975: 11) The whole situation is one of paranoid fear and dread, 'persecution anxiety' for 'the preservation of the ego' (Klein, 1935/1975: 269). This internal persecutory situation constitutes, for Klein, the traditional, Freudian superego driven by anxiety for self.

The persecutory inner world determines the sorts of narrative phantasies that can make sense, and so appear, in it: the sorts of figures that are phantasised as occupying it – persecutors and ideal figures; the sorts of things they are phantasised as doing – paranoid phantasies of dreadful threat or omnipotent power coupled with strict expectations; and the sorts of things the infant can phantasise doing in order to try to cope with or defend against this threat – the 'paranoid-schizoid defences': phantasies of omnipotent denial of the persecutory situation, in which the infant turns to and identifies with the idealised good object, which is 'equal to annihilation [of the

persecutors] by the destructive impulse.' (Klein, 1946/1975: 7) That is, in identifying with the omnipotent ideal object it launches sadistic destructive attacks on the persecutors in an 'omnipotent annihilation of the bad persecutory object and the painful situation,' (Klein 1946/1975: 7) which may involve both 'oral' 'impulse[s] to suck dry, bite up, scoop out and rob the mother's body of its good contents,' and 'anal/urethral' attacks, i.e. ideas of 'expelling dangerous substances out of the self and into the mother' (Klein 1946/1975: 8).

Because such phantasy attacks are focused on oral functions, however, they bring with them the inevitable corollary of phantasised introjection of the persecuting object, of taking in through the mouth, a bad object now (as a consequence of all the biting and chewing) fragmented into very many persecutors which thus end up (in phantasy) inside. The momentary relief thus only reinforces the internal terror, and the response to this is projective defence. The bad things now felt to be inside are phantasised as being expelled. The phantasies of expulsion are bound up with projection of bad parts of the self *into* the world, and in particular into the bad breast (the frustrating mother's body), and these are felt to persist inside her: the relief is both a felt ridding of the self of badness and a felt *control* over the bad external object because the expelled bad parts are now felt to be inside the object and are imagined to manipulate and master it, so that it might no longer be able to frustrate. This process of *projecting into* Klein calls 'projective identification' (Klein 1946/1975: 8).

The main formal characteristic of this primitive, persecutory inner world is fragmentation or '*splitting*'. Splitting is a pervasive feature of primitive phantasy, according to Klein, and is important in healthy development, but at this stage it is not so much an active *doing*, the exercise of a capacity, as the mere upshot of an *incapacity*. It is because the infant has yet to develop the capacity to *apprehend 'whole objects'* that its inner world is built up of introjected simplistic 'part objects' – very good, ideal ones and absolutely bad, dreadful persecutory ones.⁵ The persecutory inner world is characterised pervasively by this 'splitting' of objects into absolutely good and absolutely bad. This splitting effect is the result of what we might call a metaphysical inadequacy endemic to early mental life: the lack of the category of substance (Wollheim, 1984), that of a persisting whole that may be complex and that may manifest contrary qualities at different times, or even at the same time, and that has the capacity to undergo and to undertake change. At this stage, rather than apprehending things as complex and changing persisting substances, an object's identity is governed by the emotional attachment to the object: objects are categorised as one or another by their being good, i.e. loved, or bad, i.e. hated.

4. From Splitting to Wholeness – the Depressive Position

⁵ This is not the whole story, for Klein also talks of splitting at this early stage as a kind of accomplishment, and of the inability to keep good and bad objects apart as a problem. How to make sense of this I leave for another time.

At a certain point, though, a major advance begins to take place: a 'cognitive achievement... the acquisition of the concept of a whole object or a whole person' (Wollheim, 1984: 207). The move to being able to apprehend whole persons marks a radical shift in the categorial framework of experience itself: the appearance of the category of substance, of a whole that can have various, perhaps conflicting qualities, and undergo change in time. This achievement has massive ramifications, for now the infant begins to be able to put together – indeed, cannot keep apart – the good and bad objects, coming to see that they were all along really only *parts* of a complex whole object (the primary care-giver), one which persists through time and is sometimes good, satisfying, loved, and sometimes bad, frustrating, hated. And this move 'from a partial object relation to the relation to a complete object' (Klein, 1935/1975: 264) ushers in a whole new emotional situation in which 'anxiety and feelings of suffering are of a much more complex nature' (Klein, 1935/1975: 271).

In the persecutory position, anxiety is primarily for the *self*, but even the anxiety for the good objects that are also threatened by internal persecutors is anxiety about the attacks they face from what are felt to be alien persecutory figures. In other words, all anxiety is third-personal as regards the source of threat. Now, in what Klein calls the 'depressive position', a new realisation dawns: that the good, loved objects the infant has been trying to protect, and the bad persecutory figures that it has been attacking and trying to destroy in phantasy, turn out to be merely different aspects of one and the same, whole, complex object or person. Splitting begins to recede and the good and bad objects begin to be brought together or 'integrated', recognised as aspects of the same whole object. This involves the infant's recognising that its hateful and attacking attitudes to the bad objects are really directed at merely the bad aspect of the good object, and thus confronting the psychic reality of its ambivalence toward what it loves. The infant's recognition of the reality that it is their own fury that threatens, in phantasy, the loved object, brings with it 'an overwhelming feeling of guilt', not the persecutory kind of guilt, but of 'an overwhelming feeling of guilt' or real remorse toward the object, 'a sense of responsibility for preserving it intact against persecutors and the id, and with sadness relating to expectation of the impending loss of it.' (Klein, 1935/1975: 270) The fragments of the object that seemed like persecutors now seem like damaged, injured bits of an ultimately good whole that need to be put back together, rather than fundamentally bad things. This new phantasy world involves recognition of complexity: of the ambiguity of the object, and the ambivalence of the self that this entails. The challenge is fundamentally about coming to terms with psychic reality – with one's own complex and conflicted emotions, the co-existence of love and hate regarding the same object, and the sense of remorse and responsibility that comes with this.

This is a qualitatively new kind of anxiety that is not third-personal but first-personal – *depressive* anxiety, real guilt or remorse, and it 'is interwoven with continuous and desperate efforts to save the good objects.' (Klein, 1935/1975: 269) At this point, 'the ego feels impelled to make restitution for all the sadistic attacks it has launched on that object, 'and if things go well the

infant is able to embrace this, to work through the 'depressive position', through the acting out of creative and hopeful phantasies of 'making reparation to the object' (Klein, 1935/1975: 265; see Wollheim, 1984: ch. 7; Speziale-Bagliacca, 2004: ch. 4). Because objects and emotions are no longer being split into simplistic good and bad, the complex and messy reality of conflict can be apprehended, and it is at this stage that 'a full capacity for love' for a whole person becomes possible. becomes possible, because love for the idealized object was never really love for something real; it is only in working through the depressive realisation that love for real whole objects rather than false part objects can be encountered.

But things can easily go awry: the depressive anxiety may be too overpowering if the infant cannot discover hope through well-received (see Winnicott, 1954/1975) reparative phantasies, and then one of three things may happen. First, the infant may sink into a *depressive despair* in which they feel overwhelmed by the damage they have done to the good object and have no hope that they can mend and restore it, no faith in their own goodness and capacities. Second, they may resort to the *manic defence*, phantasies of omnipotent denial of depressive anxiety, of the self's vulnerability to the persecutors and of its concern for the good objects. These combine phantasies of a persecutory nature with phantasies of reparation in over-drive. The *telos* of these phantasies is '*controlling and mastering* objects' (Klein, 1935/1975: 277), 'of overcoming and humiliating it' or 'controlling it or keeping it in suspended animation.' (Klein, 1940/1975: 351) The manic defence, then, is rooted in phantasies of an unrealistic nature that deny *psychic reality*, but it is also responding to both persecutory and depressive anxieties in relation to realistic, whole objects. The infant's third option is to flee from the depressive situation altogether, regressing to splitting and paranoia of persecutory phantasy. This involves renouncing the integration of the object and the ambivalence of the self, but it leads back of course to the same persecutory anxiety situation, and the schizoid defences of the persecutory situation are then its only recourse, leading it back into the 'vicious circle' of the persecutory position (Klein, 1946/1975: 15).

5. Phantasy and Reality

Now the persecutory and depressive 'positions' are not merely developmental stages. One does not simply pass through them. Rather, they lay the foundations of the two distinct forms of unconscious phantasy world that organise a person's unconscious activity, and so conscious experience, throughout life. The developmental process is not one that is done with. The working through from the persecutory to the depressive position and through the latter is an unending task rather than a job that is ever completed. To differing extents, we are all always-working-through the depressive position. The challenges of the depressive and persecutory phantasy worlds, and the fluctuation between these, are constantly at work in ordinary, everyday experience.

Why? How do these infantile, unconscious phantasy worlds bear on conscious adult life?

The answer is that phantasy is not something that intervenes in conscious experience from the outside; it always-already structures conscious experience. As Wollheim suggests, we may think of unconscious phantasy as an unconscious drama, with its *genre*, its '*dramatis personae*' of introjected figures, each with their 'repertoire', and the consequent narrative arc. This unconscious drama is in interplay – with varying degrees of openness to to-ing-and-fro-ing – with experience of the real world and the real figures in it. Unconscious phantasy emotionally organises conscious experience in something like the way an improvisational play organises the action: it provides the overarching framework that both is developed by the performers in the performance and yet constrains that performance. It is something like a deep interpretive schema through which the significant figures in one's experienced world and one's relationships to them are experienced. Unconscious phantasy is a psychological retrieval of Kant's conception of a transcendental *a priori* framework of categories which is necessary for any experience and which shapes the raw data into a coherent object-world: it is an emotional *a priori*.⁶

But of course there are not one but two fundamental kinds of inner phantasy world with very different emotional-categorical structures. Fundamentally, persecutory phantasy, because it is so simplistically organised into split good and bad figures, is unrealistic and forces emotional experience into a false oscillation between abstract extremes: other persons can show up only as ideally good or as terribly bad. In the persecutory inner world, the *dramatis personae* consists mainly of dreadful persecutors, their repertoires essentially that of sadistic and dreadful attacks on the self, along with idealized good objects. Because of the splitting at work, these are ideally pure and good, and the mass of persecutors have had any goodness stripped out of them: they are *all bad*. In this split world oscillations between the extremes are typical, because persecutory phantasy has no room for complex wholes, for the in-between, for the difficult and messy and complicated, because it has no room for ambivalence. Its defences are paranoid-schizoid omnipotent, sadistic ones: there is no space here for creative, reparative solutions here, because there is no complexity to begin with. Since the persecutors are absolutely bad figures and anxiety is consequently primarily for the self, omnipotent destructive attacks are all that make sense. Depressive phantasy allows for complex and realistic figures to enter into experience, and thus for complex attitudes toward them in too – ambivalence, remorse and responsibility – and in turn for creative responses of engagement, reparative and rebuilding activity through which hope, tolerance of ambivalence and love can be strengthened, and deep and lasting sense of security may be achieved.

6. A Kleinian Account of What Punishment Expresses

6 The failure to appreciate the categorical nature of unconscious phantasy, its functioning as an *a priori* framework of experience, can lead to incredulous incomprehension of psychoanalytic concepts like omnipotent denial (e.g. Cohen, 2001)

Needless to say, I think Klein's is a plausible and powerful account, and its influence in post-Freudian psychoanalysis is testimony to its force. I hope my brief reconstruction of Klein's position shows some of its plausibility and might forestall misunderstandings, but beyond this I shan't try to directly defend her account here (I refer the reader to the philosophical defences of Klein's position mounted by Wollheim (1984) and Gardner (1993)). Instead, I want to explore the significance of Klein's account, supposing it were true, for expressivism about punishment. The account's fruitfulness here offers some indirect support for the theory, I think, for insofar as the account is illuminating this lends support to the contention that it is true. Now we saw that philosophical accounts ignore completely, while psychosocial accounts typically abstract from, the significance of unconscious phantasy, and thus the holistic unifying and organising role that the unconscious inner world plays in shaping experience (for examples of deployments of Klein in thinking about crime and punishment, see Hinshelwood, 1995; Hollway and Jefferson, 2000; Valier, 2000; Brown, 2003). My basic claim is that what punishment expresses is, formally, not this or that emotion or defence, but rather *unconscious phantasy*, which organises the constellation of emotional reactions and defences. Punishment expresses the drama of the inner world of unconscious phantasy mediated by the symbolic machinery of the political. By making unconscious phantasy central to interpretation we can both make intelligible the unities of elements that we find in penalty, and can comprehend the antinomies and oscillations between those unities that are characteristic features of the penal-historical landscape.

I want to support this general claim with three more substantive, tentative and illustrative theses. First, most basically, punishment in modernity expresses the *persecutory* inner world, and getting this fact into view allows us to locate many central features of modern punishment within a framework that makes them intelligible in their unity. Second, in addition to the dominant persecutory phantasy structure of modern punishment, we can interpret the historical development of modern punishment as playing out repeated unsuccessful attempts to move beyond persecutory phantasy by confronting and working through *depressive* realisations and anxieties. Third, the prominent concrete forms of punishment can fruitfully be interpreted in terms of intrinsic symbolic fit as playing out narrative elements and defences belonging to the persecutory and depressive moments within this framework. In penal history we can trace the playing out in political terms of the failed attempt of liberal modernity to work through the depressive position, and the consequent regression to persecutory phantasy. We find in other words the political problems of modernity being organised and distorted by competing emotional *a priori* structures of unconscious phantasy, and in particular by the inability to encounter, tolerate and work through the depressive moment: the repeated failure to face and deal with a realistic apprehension of both the ambiguity and the ambivalence of liberal modernity's deeply conflicted damaging aspects. I'll sketch some of the contours of these theses in only the broadest strokes before considering some further interpretive

advantages of the Kleinian lens and then reflecting on its implications for reconciling or integrating psychosocial and philosophical expressivism.

7. Persecutory Phantasy and Modern Punishment

In the founding modern philosophical images of social order, the basic feature of the political world is *splitting* – the pervasive separation of good and bad, vulnerable and dangerous, and the corresponding unleashing of simplified, one-sided emotional attitudes of love and hate, so that the political world is made over in the image of the infantile persecutory drama. In Hobbes we find these two images of the subject and of the state side by side uneasily, with no real acknowledgement of the tension (Hobbes, 1651/1991), but in Kant this splitting of political world turns metaphysical: now the subject-citizen is conceived as a strange amalgam of intelligible reason and empirical nature (Kant, 1929: A532/B560-A558/B586), where reason is free, moral, good, and nature, sensuous need, is equated with badness, danger, evil (Kant, 1993: Pt. Bk. 1). It is because of sensuous nature in the subject that the autonomy of the citizen is vulnerable: acting on natural inclination is, for Kant, the source of radical evil, and that other citizens may act on inclination is what makes them a danger to my autonomy. Punishment is in one sense the primary institutional performance in which the various coping responses to this terrifying and threatening world of dangerous persecutors that underlies the liberal political image are expressed. The basic feature of the penal narrative is the splitting-off of the bad from the good, the separating out of the dreadful persecutors, phantasies of purifying the good and destroying, getting rid of or controlling the bad.

Of course, the logic of retributive desert represented in some sense a curtailment, a sublimation, of persecutory phantasy, yet it remained rooted in it: though a civilisation of revenge, it remained rooted in the phantastic model of retaliation that is central to the persecutory. How are we to understand this logic of retaliation? In Klein, it is explained by projection. The infant wants to retaliate to the frustrating bad object with attacks because it has projected its own rage onto the object and so now sees it as thoroughly bad and threatening. But in turn it anticipates the object's retaliation, because of the projection of its retaliatory impulses onto the object. Rage and retaliation *make sense*, because the object is not experienced as a complex whole that also has good aspects, that has been damaged by the self, and that could be engaged with as a whole person rather than being simply destroyed or controlled. This explains the fusion of repressive attitudes to the self with strict moralistic demands on others and eager retributive impulses toward them, such as we see in the classical Kantian ethic. The intimate connection of strict moralism with retributive lust belongs to the persecutory superego (see Money-Kyrle, 1955).

We can identify each of the three main defensive phantasies of the persecutory position expressed in punishment, in varying degrees at different points in modern punishment's history. In

the early modern period, what Foucault calls the 'spectacular' and 'carcereal' is prominent (leaving aside actual incidence) in the public image of punishment: the penal performance is dominated by the public and visible practices of flaying, burning, tearing and dismembering of the offender's body (Foucault, 1977; Spierenburg, 1984). Such spectacular punishments belong, symbolically, to the 'oral-sadistic' annihilative attacks on the mother's body in primitive infantile phantasy: omnipotent denial of the threat and omnipotent destruction of it are brought together in the phantastic identification with the idealised punishing state that in annihilating persecutors both affirms and denies the danger: the persecutors are indeed everywhere, but the state will omnipotently destroy them. In this period it is no coincidence that the identification with the omnipotent ideal object – the state – was fragile, just as it is for the infant in the persecutory position: the state stands for the idealised good object, and brings together 'the very strict demands' and the 'cruelty' of the ideal good object in the persecutory inner world. The state conceived in this way can, because of its strictness and violence, easily become a persecutor of the self, or citizen, if its demands are not adhered to, if its wrath cannot be placated: oscillation from ideal to persecuting object is a constant danger. The actual spectacular violent attacks upon the offender's body have the symbolic significance of phantastically destroying the persecutors felt to be everywhere in the political community that is merely superimposed upon a paranoid state of nature. But this phantastic destruction of persecutors is of course unsustainable.

In the transitional period that follows, the rise of transportation, for example, as a punitive strategy, the emphasis shifts from spectacular annihilation of the persecutor (offender) to the expulsion, the getting rid of, bad objects: transportation (leaving aside its economic function) can be interpreted symbolically as enacting the projective expulsion of bad objects to which the infant resorts, an expulsion that brings with it the benefit not available to the infant – that the bad objects are sent away to a far off land never to return, never to be introjectively 'taken-back'. But in the transition to what Foucault calls the disciplinary, incarcerated regimes of punishment, expulsion of bad objects is now carried out by separating them off into prisons conceived as civic purgatory where citizenship is suspended, but now the benefit of transportation is given up and the infantile situation is more accurately re-enacted: the projectively expelled bad objects must ultimately be reintroduced, reincorporated, taken-back into the community.

Through all this, punishment expresses a phantasy of expulsion of bad parts of the political community into the offender, and in this punishment expresses the disavowal of the community's aggressive and sadistic impulses – evident in the indifferent perpetuation of deprivation, degradation and social exclusion – by projecting them into the offender. The illusion achieved is analogous to that achieved by infantile projection: what the community and its citizens shield themselves from in retaliatory attacks on the offender is the recognition that it is the structural violence and mutual hostility that are the organising principles of liberal capitalist societies that have made the offender into the persecutor in the first place.

The social failures and structural violence of our community which generate the circumstances in which serious offenders are forged may be interpreted as acted-out projections by the community of its badness and hostility onto the person, which turns them into a persecuting figure in turn.⁷ The problem is that the projection or ejection of bad figures out of the community – and into incarceration – is to put them into a place where renewed attacks are made on them, further damage done, and this raises further anxieties about the inevitable taking back in of bad things that have been further spoiled. Infantile phantasies of spoiling the object but then having to reincorporate it, damaged and poisonous, are elaborated in the parole mechanisms of risk assessment and the contemporary machinery of post-incarceration monitoring. The underlying phantasy narrative here is one of fear at the prospect of taking in something which one has taken in and made bad, ejected and made worse, and must then take back.

The disciplinary model manifests the second and more complex form of persecutory defence – the expulsion of bad aspects into the persecutor as a means not so much of harming it as of entering into and taking control of – mastering and manipulating – it. This projective identification 'in some ways... implies a deflection of guilt from the self on to the [other] person,' (Klein, 1946/1975: 12) or from the community onto the offender, insofar as it is a splitting-off of the bad aspect of the community that is felt to be a danger to it and to all that is good in it – including perhaps the community itself's failures that create dangerous offenders – and the projection of this badness into the offender themselves as if it arose *sui generis* in them. But this projection entails also a controlling phantasy of overriding the independence and thus the danger of the persecuting figure by manipulating them from the inside. There is no better exemplification of this phantasy than those sociobiological dreams of a future in which 'parolees will check in and be monitored by transmitters embedded in their flesh, reporting their whereabouts in code in code and automatically,' or in which 'two-way sub-dermal' computer chips could be used to generate a 'transponder surveillance system [surrounding] the criminal with a kind of externalized conscience' (Meyer, 1971. See generally Delgado, 1969).

The recent 'preventive turn' (Carvalho, 2017; Ashworth and Zedner, 2014) has attempted to get as close to this ideal as possible with its web of anticipatory control and monitoring apparatuses that have been expanded into the fabric of civil society. As projective identification (see Braddock, ms.) is used psychologically to rid the citizen and the community of bad elements, and this in turn produces a certain controlling power over the person into whom they are projected, so the offender is in practice forced to take on the badness projected into them by the state in the punishment performance: they are expected to admit they are bad, not to minimise their offending, etc.; on the other hand, prison often makes people worse, and the whole punishment process risks

⁷ There is a lot to unpack here, and no room, but the thought would be that social violence perpetuated by our political communities is explicable as the projective disavowal of our own vulnerability. Disdain toward the disadvantaged, motivated by a desire to distance oneself from it, is a well-documented tendency in the cognitive bias literature on belief in a just world and fundamental attribution error (Lerner and Simmons, 1966; Bertram 2006; Jost et al, 2004).

encouraging offenders to take on an identification or 'label' as bad (Becker, 1963; Lemert, 1951), and this can in turn lead to their internalising the projected badness and becoming worse, acting out the bad identification (Maruna, 2001; Schlosser, 2015).

Paranoid anxiety of a multitude of persecutors, the pervasive splitting of the world into simplistic good and bad figures, the projection of hostility and violence onto split off bad figures, and omnipotent sadistic phantasies of retaliating by destroying, spoiling, taking over and controlling, the persecutors, coalesce around the underlying dramatic situation of the persecutory phantasy world. This may help to explain the overt irrationality and irrationalism of many features of modern penalty while allowing us to make a kind of sense of them by seeing how they fit together into a drama that though not rational in the conscious, realistic sense nevertheless has a certain immanent logic and coherence of its own. Punishment plays out politically the paranoid and split phantasy of a persecutory, hostile world that has always been at the core of liberal modernity's self-image.

8. Depressive Despair and the Persecutory Return

If this is right it raises a question. To rehearse a well-worn and probably oversimplified story, classical retributivism went on the wane in the late 19th century, and throughout much of the 20th century it was displaced, seemingly for good, by a medical or pathology model of crime and a therapy or treatment model of punishment, only for this trend to be sharply reversed in the 1970s and 80s with the return of retributivism (the 'just deserts' model) in theory and practice (i.e. Von Hirsch, 1976; Morris, 1976; Murphy, 1979; see Hudson, 1987: chs. 1-2). There seems to be a century-long lull in the dominance of the persecutory shape of punishment. How is this to be interpreted on a Kleinian view? My suggestion is that what we see in this historical story is precisely the 'fluctuation' between persecutory and depressive phantasy that for Klein is endemic to the living of a human life. The rise of the 'treatment' model can be understood as a manifestation of liberal modernity's encounter with the depressive position. The turn to a 'treatment model' of punishment was part of a wider political change in which the structurally violent and damaging nature of the political community was reluctantly encountered, spear-headed at first by religious and philanthropic groups and the emerging labour movement, and then by a project to address liberal capitalism's systemic failures undertaken, with the social housing, basic educational provision and workers rights, the beginnings of what would become the new political settlement of the welfare state and a transformed notion of citizenship, and the treatment model of punishment rises with this emerging transformation.

This is something like a 'metaphysical advance' by which it becomes possible for liberal politics to move beyond simplistic splitting that idealises the community and state, and demonises offenders as persecutors to be sadistically or paranoically dealt with. Now, the offender-citizen

comes to be recognised as a whole person, an ambiguous figure who is *both* a wrongdoer and potentially harmful *and* also a damaged but potentially good figure. The wrongful harm they have done is now recognised not so much as the manifestation of inherent malevolence as the upshot of the wrongful harm they have suffered at the community's hands. The world is no longer organised by a split between rational beings and natural animals; we become able to tentatively to apprehend one another as *vulnerable rational animals*.

Things move beyond the persecutory anxiety situation in this advance, but a new – depressive – anxiety situation arises: recognising the ambiguity of the offender and the harm done to them entails recognising the political community's own ambiguity – its real ambivalence, and especially its harmful and damaging structural attacks on many of its members. This is just the ambivalence of a capitalist civil society that Hegel had recognised without being able to process, and Marx had driven home more convincingly: the tension between a liberal ideal of equal liberty and the social actuality of poverty, exploitation, oppression. Crime and retributive punishment now both came to be seen as an after effect of a 'wrong done to one class by another' (Hegel, 1821/1942: 278), as merely the perpetuation of the vicious cycle of sadistic persecutory attacks, or as 'particular injustice' in response to 'injustice in general' (Marx, 1844/2000: 81). Such recognition of the community's own ambiguity and ambivalence provokes depressive anxiety, genuine guilt or remorse, for the community's harmfulness, a topic which became thematised as a matter of *social justice*, and thus a move away from persecutory anxiety for the community's well-being in the face of the persecution of the simplistically conceived offender, to anxiety for the citizens' well-being, including those who become offenders. It marks a move *from civic splitting to civic integration*. For offenders too are now recognised to be *loved whole citizens* whom the community has been harming, and who have become harmful to it in turn, a thought which is less strange than it sounds once we put it into the discursive context of *civic love* (see Vlastos, 1969/1981).

The anxiety of this depressive realisation, we might think, proved too much for liberal modernity. At first, the response is what we would expect: a *drive to make reparation*, to repair the damage and restore the object, both more widely in mending the damaging structural inadequacies through the expansion of the social security infrastructure, and within punishment through the rise of the rehabilitative aim, of punishment conceived as a kind of treatment to restore the damaged citizen. In the post-war period when this project was carried through we see the treatment model becoming the orthodoxy (see e.g. Wootton, 1959), while retributive punitivity came widely to be regarded as an irrational throw-back to less civilised times. But because such efforts were too superficial and limited they were largely unsuccessful, and hope was difficult to sustain.

The story told by the new retributivists was that the treatment model had really been a *reifying* one that failed to respect the offender as a person, instead treating them as an object to be manipulated (Strawson, 1962; Morris, 1976; Moore, 1996). There has always seemed something fishy in the assertion that only by punitivity do we respect people, but there is a grain of truth in the

more general diagnosis, and we can now see what: reparative efforts, when disappointed, give way to an oscillation between, on the one hand, depressive despair, and on the other, a manic defence against this hopelessness – overactive omnipotent efforts at reparation that are bound up with phantasies of control and mastery. The manic defence involves elements of self-concerned persecutory anxiety about what should happen if the damaged objects cannot be healed and repaired and should then persist inside the self as dangerous contaminants, as well as depressive anxiety concerning the fate of the objects themselves, which sums up pretty well the ambivalence of the rehabilitative ethos. The manic phantasy response, though, is one of omnipotent power over the objects so that there is both an unrealistic faith in the power to repair even though this is unrealistic, detached from evidence, and, through the way this reparation is carried out, an omnipotent objectifying thinking that imagines the objects to be *mere* objects under the complete control of the self.

To the extent that punishment on the treatment model was in practice reifying, this may be interpreted not as an inevitable tendency of reparative impulses but as the consequence of the controlling impulses of the manic defence against depressive realisations. Inevitably, this defence eventually failed. The retributive revival of the 1970s and 80s was preceded by a depressive moment – the sense that 'nothing works' and that crime is inevitable, i.e. that the community is not good enough to avoid or to repair the damage, trauma, poverty, deprivation and exclusion that makes offenders.⁸

On the other hand, the retributivist revival has turned out to be just as concerned as the treatment model with obsessive control, and if retributive punishment expresses a split phantasy world of part-objects, it may be retributivism that is guilty of reifying: of seeing persons as non-persons, mere bad parts lacking the wholeness and complexity essential to personhood. For the regress to the persecutory world, although avoiding the despair of depressive anxiety, reactivates the primitive self-concerned persecutory anxiety and brings back the primitive defences against it, plunging the community into a paranoid world that is still obsessively controlling but has abandoned the recognition of its own ambivalence along with the desire to repair things. Indeed, we might with some justification describe the punitive and preventive reorientations of the recent neoliberal period of punishment and crime control (see Ramsay, 2012; Ashworth and Zedner, 2014; Carvalho, 2017) as a *persecutory turn*, or rather, *return*. Punitive and preventive criminal law consumed by paranoid phantasies of an insecure world is continuous with the persecutory phantasy world in which modern punishment was rooted from the start: a persecutory phantasy world dominated by splitting, projection and introjection and characterised above all by paranoid fear and destructive wishes against persecutors felt to be everywhere. The obsession with

8 I have centrally in mind serious and persistent offenders engaging in *mala in se* crimes of violation. Obviously, not all crime is traceable to such factors – a significant proportion (e.g. drug possession offences) is arguably traceable to nothing other than inappropriate criminalisation.

omnipotent control that is entwined with persecutory retaliatory phantasies helps to make sense of the present predicament in which retributive punitiveness and omnipotent control are central symbolic elements of punishment.

The story of punishment's expression is that of liberal modernity's failed attempt to cope with its own ambivalence between liberty and equal respect and hostility and structural violence. Punishment's changing narrative surfaces play out its abortive attempts to work through depressive realisations to the achievement of civic love.

9. The Psychosocial Significance of a Kleinian View

More generally, where does a Kleinian view get us? First, it avoids eclecticism. Rather than postulating a variety of seemingly disparate elements of punishment's expressive content, it allows us to make sense of the connection of such aspects as integral elements of an unconscious phantasy structure, or to separate them as belonging to distinct phantasy structures. Consider, for example, sadism: the sadistic aspects of punishment are often interpreted *via* Nietzsche's postulate that underwriting the desire to harm or destroy is a more rock-bottom, natural sadistic pleasure in causing suffering *per se*. Now there is no doubt that the eroticisation of the causing of another's suffering is a real psychological possibility, but the Nietzschean view needlessly gives such an impulse rock-bottom status as a basic drive in our nature. Klein shows us that 'sadistic' urges and phantasies – those involving satisfaction or relief from the harming or destruction of the object – can be understood without positing a rock-bottom sadistic urge. For these make sense in the wider context of persecutory phantasy and the paranoid anxiety about being destroyed that it structures. Many of the sadistic aspects we find in persecutory phantasy are concerned with preservation of the self from dangerous persecutors rather than pleasure in cruelty *per se*, and the gratification of destructive attacks may be amply comprehensible in terms of the mere need for relief from crushing fear.⁹ In other words, it is plausible that sadistic phantasies are evidence not of a basic drive to cause suffering but of a basic drive to avoid anxiety.

If Nietzsche is right then the lust for cruelty is basic to human beings and there is no real hope of being rid of sadism, but if Klein is right then the desire to harm is dependant on the fear of being harmed that pervades a paranoid phantasy world, and there is hope that we can move beyond it. Analyses of punishment that focus on the idea of a sadistic lust for cruelty would then be off the mark, even where punishment takes overtly sadistic forms, in overlooking the basis of such impulses in an underlying persecutory phantasy world. The carnival of cruelty is something that

9 This is of course to gloss over complications. Klein held to late Freud's hypothesis of a death drive as fundamental, alongside eros, in human nature, which may seem to bring her closer to Nietzsche. But the death drive for her is not primordially a sadistic disposition to take pleasure in causing another's suffering, but rather an 'instinct ultimately directed against the self' (Klein, 1948/1975: 29), a *disintegrative* tendency, relief from which is, seemingly, achieved by redirecting it onto others.

can get traction only insofar as one's experience is structured by the projection and splitting that make possible a world of simplified good and bad.

A different example is guilt, to which we find frequent reference in the literature, but in conflicting ways. Garland illustrates the point well. In *Punishment and Modern Society* he writes of 'punishers from a sense of guilt', whereby

an unconsciously punitive attitude towards one's own anti-social wishes may carry over into a projected punitive attitude towards those who have actually acted out such prohibited desires. (Garland, 1990: 240)

In later work Garland alludes again to guilt expressed in punishment:

These middle class fears [of crime] were, no doubt, overlaid by an element of guilt and bad conscience. This was, after all, the society that these classes had chosen... [in which] the delicate web of solidarity and community [had been undone]. (Garland, 2001: 156)

In the first passage, the guilt at issue is that of the persecutory superego menaced by strict demands against the id accompanied by fear of retaliation, and its 'carrying over' to a punitive attitude toward offenders seems best interpreted as involving projective identification, whereby the self envies the offender, and identifies with them by projecting into them their harmful wishes, so that their guilt – the internal persecutors' attacks on the self – can be directed outward at the offender. This sort of guilt, along with its disavowal, belongs to a persecutory mindset.¹⁰ But in the second passage, a quite different kind of guilt is in play: remorse for the harm and damage one has caused to the object. This is the depressive guilt that liberal modernity has not been able to work through, but which it is constantly confronted with. The response to depressive guilt *can* be creative and reparative activity that builds hope in oneself and trust in the object. This seems to have proved too much, leading to manic and persecutory defences against despair instead. The immediate defence against such guilt is manic omnipotent control, but such phantasies, in which desire for 'triumphing' over the object, are intermingled with anxiety about 'humiliating' it and thus of provoking it to further persecuting, retaliatory attacks. Such defences quickly degenerate into persecutory fears and to the resort to paranoid defences, phantasies of destroying the object, so that the manic and paranoid defences operate in oscillation. The punitive and preventive practices

¹⁰ It should be noted that to talk of 'persecutory guilt' is vaguely controversial. It is often said that Klein distinguished between persecutory *anxiety* and depressive *guilt*, but she clearly understands persecutory anxiety to underpin one form of what is called 'guilt' – the ascetic form of guilt which Nietzsche and Freud take as the central case. She tends to call this persecutory 'anxiety' rather than 'guilt' to contrast it with what she sees as *real* guilt, remorse (see Klein, 1948/1975), but it is clear that she takes persecutory anxiety to explain the ascetic form of guilt from the ways in which she identifies persecutory anxiety with the 'overpowering strictness in the superego' that is typical of ascetic guilt and conscience (Klein, 1934/1975: 258).

of the persecutory turn defend against depressive guilt by regression to the persecutory world of splitting and the defences of manic control and paranoid destructiveness that are available in it.

Second, a Kleinian view gives us the conceptual resources to make sense of both unity and antinomy in penal expression. We can locate the various elements of modern penal expression within underlying unities – the persecutory and the depressive – that are categorially distinct and incompatible structures, and which thus exist in and give rise to antinomy, oscillation between contradictory and incompatible perspectives. For these unconscious *a priori*s involve not only different kinds of defence against anxiety, but different kinds of anxiety against which defence is sought.

The dominant paradigm is persecutory phantasy, both at the origins of liberal penalty and with renewed vigour in the hostility and paranoia of the current punitive-preventive phase. The central anxieties are of a persecutory, paranoid nature, that is, concerned primarily about the danger of destruction or disintegration of the self, i.e. of the subject and the community with which they are identified. 'Ontological insecurity' is the order of the day (Giddens, 1991; Bauman, 2000; Ramsay, 2012), a pervasive sense of fear felt to come from persecutors perceived to be invasive and everywhere. This unrealistic experience of the social world involves disavowal of social badness through the primitive persecutory defences – projection and splitting – which among other things conceal the fact that the originary hostility that has formed the world of persecutors – both real and phantasised – has its source in the community and its members. And the hostility of the world thus coloured by persecutory phantasy provokes hostile and controlling responses, including the more sophisticated persecutory defence of projective identification, projecting into, not least because such simplistically bad persecuting figures seem beyond any other form of response – they seem to be monsters to whom the only proper response is obliteration or manipulation.

Depressive realisations involve a more realistic view of the citizen (including the offender) and the community, as whole complex objects rather than abstractly good or bad part objects. But this recognition of complexity and ambiguity entails recognition of one's own – the community's and the citizen's – ambivalence, and hence provokes the depressive anxieties of remorse and self-questioning. These anxieties are concerned primarily not with the preservation of the self or the community, but with concern for the preservation of the object, and particularly its goodness, in particular in light of the damage done to it by the community. Depressive anxiety is ultimately concerned with the community's own goodness, rather than its mere preservation: the anxious question is not so much about our preservation but about our goodness. Here a sense of responsibility for the community's role, its damaging nature, and recognition of the community's own ambivalent nature, motivate attempts to view offenders with understanding and compassion, and to provide for reparative and therapeutic mending on both sides, through the provision of therapeutic support and of opportunities for the offender to make meaningful reparative efforts and recover hope and trust (see Lacey and Pickard, 2015). If reparative activity cannot sustain hope in

the goodness of the community, descent into despair, flight into mania, and ultimately regression into persecutory phantasy – which essentially represses ambivalence – are the available moves.

Of course, in the persecutory turn depressive, reparative impulses have not disappeared from punishment – in fact, there have been significant advances in the development of such therapeutic approaches within punishment, including within prisons and offender programmes (see Maruna, 2001; Stevens, 2012; Lacey and Pickard, 2015; but c.f. Schlosser, 2015). But this paradigm exists in a persistent antinomy with the dominant persecutory one. As Ethell puts it, the more ambivalent reactive attitudes of the depressive moment bring with them 'an awareness of our vulnerability to circumstance, of the fragility of our moral character,' (Ethell, 2010: 254) and thus our common personhood as beings vulnerable to the damaging conditions of a community. It involves recognising the ways in which we are shaped by our past and our context, and the ways in which the community is complicit in the harming of some of its members in ways that can lead to them becoming people who violate others. Thus, depressive anxiety, the recognition of ambivalence, remorse and responsibility on the part of the community, are inevitable so long as this viewpoint is occupied, which might explain why it remains unstable while persecutory, retributive phantasy is dominant. The tendency to regress to persecutory splitting instead of facing depressive anxiety is pervasive in modern penalty.

11. From Psychosocial to Philosophical Expressivism

Such a view can inform philosophical expressivism in several ways. First, as we have already seen, it allows us to trace and locate the intrinsic symbolic fit between penal practices – 'our symbols of infamy' – and underlying phantasy structures, thus avoiding conventionalism. Second, the capacity to make sense of the unities and antinomies in penal expression has philosophical significance for understanding what have often seemed to be irreconcilable and mutually incomprehensible intuitions in the history of the philosophy of punishment. The basic incompatibility between the persecutory and depressive inner worlds, the insoluble antinomical conflict between paranoid, punitive, and depressive, reparative, intuitions, may help to make sense of the basically 'antinomical' character of the philosophy of punishment (Norrie, 1991; 2000).

Such antinomical oscillation between persecutory impulses toward part-objects and depressive impulses toward whole objects is not only evident in the historic dispute between retributivism and consequentialism, but is arguably internal to retributivism itself from its inception. Kant's commitment, for example, to retributive punitivity sits in uneasy tension with the essence of the categorical imperative – the requirement to recognise and respond to other rational animals as ends in themselves and never to treat them merely as a means (Kant 1997: 41) – for, as Merle has argued (2009: ch. 3), this principle appears to rule out the destructive and controlling compartments of retributive punishment. Kant's thought is structured by a basic splitting of the

subject into a free will and a determined nature, which allows the wrongdoer to be judged guilty and meriting retribution despite their actual vulnerability to biography and circumstance as a sensuous creature of determinative inclination (Kant, 1929: A532/B560-A558/B586; see Norrie, 1991: ch. 3). But this seems to block precisely the sort of loving, depressive comportment to the rational animal as a *whole object* that the categorical imperative demands (see Rodl, ms.), propping up persecutory attitudes that deny the person's wholeness and vulnerability instead.

A grasp of the persecutory phantasy roots of such splitting helps both to elucidate what is, from a Kantian point of view, wrong with them, and to explain the psychological pull of a seemingly 'lifeless fetish' (Rashdall, 1900: 197) and the corresponding difficulty of getting into and holding in view depressive whole object experience. Equally, it may help elucidate recent communicative accounts, in which the depressive longing for reparative and reconciliatory dialogue is at the fore (see Duff, 2001; Lee, 2017), but remains ambiguously disentangled from the persecutory impulse for coercive hard treatment that undermines the ideal of dialogue and grasps for a kind of control over the Other that is incompatible with communication (see Matravers, 2011; Carvalho, 2017: ch. 6).

Third, it offers a psychosocial interpretation of what punishment expresses that avoids the psychological naivety typical of philosophical accounts and the sceptical implications of reductive psychosocial accounts. It brings to philosophical expressivism the needed moral – but neither reductive nor moralizing – psychology that makes possible the coherent, non-question-begging derivation of normative implications. A Kleinian moral psychology implies a normative critique of persecutory punitiveness without threatening global reductionism, because it distinguishes between persecutory and depressive phantasy in a non-reductive yet non-moralizing way that nevertheless has normative implications: persecutory phantasy is worse, depressive phantasy better, on criteria that are independent of our prior normative attitudes to the content or effects of either.

The first criterion is *rational health* – i.e. the capacity for a realistic apprehension of, and thus the capacity to distinguish between, external and internal (emotional) reality (see Harcourt, 2017). For persecutory phantasy has at its core mental operations that keep the person at a distance from a realistic experience of their own emotional life and, in turn, of the world. It is based on an inability to apprehend whole objects, to tolerate one's own emotional ambivalence and to accept and integrate the complexity of the world. Persecutory anxiety is not always illusory, but a situation in which it is dominant is structured by the muddling up and distortion of inner and outer reality. The depressive realisation involves a more realistic recognition of both worldly and emotional reality, and depressive anxiety is based on a fuller appreciation of the wholeness and ambiguity of valued persons, and of the complexity and ambivalence of one's feelings towards them. (It is a more complex question whether there are independent, non-circular grounds for normatively privileging any particular response to depressive anxiety: has someone who has

successfully worked through the depressive position to a capacity to tolerate depressive anxiety with hope and stability necessarily more realistic than resignation to depressive despair in the face of the tragedy of one's own and others' ambivalence and the consequent messiness of our attachments? The criterion of realistic experience here becomes less stable, for the capacity for hope in one's capacities is partly creative of the capacities in which one has hope.)

The second criterion, connected to the first and immanent to the concept of expression itself, is *truthfulness*. Expression of unrealistic persecutory experience is also liable to reinforce its unrealism, both by expressing a distorted image of the world and by suppressing something crucial, serving *denial*; or by, in the very activity of expression, discarding and disavowing the thing expressed – the *projective identification* we have already encountered, in which the artist 'repeats the losing of a piece of self-knowledge.' (Wollheim, 1993: 11-12) At different points in the history of penalty, and with the prevalence of different penal practices, what punishment has expressed and what it has, in doing so, suppressed, denied or discarded in our self-knowledge has varied. But certainly among the pieces of self-knowledge that punishment today performatively discards must be that we have become, that our community is, in a deep sense anti-social, founded on hostile competitiveness – that, as Adorno puts it, 'the more someone has espoused the cause of his own aggression, the more perfectly he represents the repressive principle of [our] society' (1974: 45).

The judgment, then, that persecutory phantasy is bad is not based on the observation that it involves hostility, fear and aggression, along with the prior evaluative judgment that such things are bad. After all, it is not clear that they *are* always bad. Kleinian moral psychology furnishes independent grounds for the evaluative judgment that persecutory phantasy and its upshots are bad, rather than simply assuming it. Persecutory phantasy is bad because it necessarily involves an unrealistic, split apprehension of both self and world. To the extent that punishment expresses persecutory phantasy, it is bad because it is unrealistic and untruthful: it expresses and reinforces rational ill-health; and it misrepresents us to ourselves, serving as a vehicle of disavowal and suppression of self-knowledge. A public, institutionalised ethical practice that expressed the working through of the depressive moment (I leave the question what this might look like for another time) would be good, in turn, not because of some prior evaluative judgment that reparative impulses are ethically better than persecutory, punitive ones, but because it would be more realistic and more truthful. Less punitive and more reparative impulses are to be preferred because they are embedded in a phantasy inner world that is organised to accommodate ambiguity and ambivalence, complexity, wholeness, and as a consequence is relatively speaking infinitely open to the world.

We may, of course, already think hostile punitivity bad and constructive, reparative impulses good, but Kleinian psychology doesn't *presuppose* such normative commitments. Rather, it implies a non-question-begging explanatory vindication of them, providing an account of the deep structure of such impulses that allows us to make sense of, deepen and redeem our pre-theoretical

evaluative attitudes towards them in light of a realistic moral psychology. For it suggests that destructive and controlling punitivity is rooted in privations of rational health, and so it is no surprise if we should feel there is something wrong with them. And for those who intuitively favour punitivity, Kleinian psychology presents a challenge, suggesting that such intuitions are rooted in cognitive distortion.

None of this is to suggest that there is no proper place for recognition that the offender has done something seriously wrong, has violated, nor to suggest that anger has no place here. Rather, it is to suggest that there are fundamentally different forms of anger and ways of recognising and responding to violation. One possibility is *persecutory anger*, whose categorial frame is a phantasy world part-objects and has aggressive, destructive wishes towards abstractly bad persecutors at its core. A different possibility is *depressive anger*, whose intentional object is a whole person, and which is animated by hope as well as hurt and disappointment, and by undefensive self-reflection rather than disavowal. While persecutory anger dismisses the part-object and aims for their suffering, control or destruction, depressive anger incorporates recognition of the whole object's value, and is animated by a longing for that person to appreciate the harm they've been doing, to want to make amends, and to want to do things differently in future, a longing of the sort Lacey and Pickard (2015) have recently advocated. Evolutionary psychology has identified these two distinct possibilities (Petersen et al, 2010; McCullough et al, 2013), but a psychoanalytic moral psychology is necessary to elucidate their deep structure and to understand both the seductiveness of the persecutory and the deep-running resistance to depressive realisations.¹¹

Now it may be objected that my account implies that much of what punishment expresses is psychotic in nature, and thus that it 'pathologizes' retributive penalty. But it is begging the question to suggest that, simply because a position 'pathologizes' certain phenomena, it is doing so illicitly. The question is substantive: are they in fact pathological? Are there grounds for thinking so that are not derivative of prior contentious normative commitments regarding them? The normative judgments that a Kleinian psychology recommends meet this test. The normative distinction between persecutory and depressive phantasy is grounded in standards that are free-standing – rational health, realistic apprehension, truthfulness. They concern, essentially, the extent to which a person is capable of experiencing the world and their own mind objectively. That is hardly an arbitrary standard: it is an aspiration immanent to self-conscious agency.

In any case, it is a Kleinian contention that psychotic forms of mental functioning are actually operative in ordinary life, in normal people, generally. It is only when they come to dominate more or less exclusively that they become 'pathological' in the strong sense of

¹¹ Lacey and Pickard (2015) have pursued the evolutionary psychology path, but such an account seems to have little to say about why one rather than another emotional response might be prevalent, seemingly bottoming-out in a decisionistic notion of our capacity to choose which one we prefer.

diagnosable as disorders. Life is an ongoing struggle to overcome persecutory phantasy and work through the depressive position, but that is not a task that is ever completed; like living a life, it is an activity with no end, although it must come to one. When the normality of the pathological is appreciated, the charge of 'pathologizing' loses whatever force it had.

At its core, the persecutory turn manifests the community's disavowal of responsibility for its hostile and damaging aspects, and a regression to a simpler, albeit unrealistic, way of carving up and dealing with the world instead that defends against depressive anxiety. Now it might be suspected that I am trying to prioritise psychology and side-line politics, but not at all. The historical story I've outlined could hardly itself be either exclusively psychologically or exclusively politically explained, and although psychoanalysis has sometimes been insufficiently attentive to the social and historical mediation of the individual, a Kleinian account can, I suggest, be properly integrated into a serious socio-historical political account. After all, the struggle between persecutory and depressive phantasy hardly goes on in a vacuum: for Klein the internal and external worlds are in constant interplay. And Klein makes the point that external events will often, even always, be crucial in understanding psychological moves. Indeed, the whole point of Klein's psychology is that the basic phantasy structures laid down in infancy are liable to be reactivated by all sorts of real events in life. In particular, when the external world really becomes more persecutory, that is liable to activate and accentuate persecutory phantasy. Thus, say, a Meadian view of penalty in capitalist society as expressing 'hostile solidarity' (Carvalho and Chamberlen, 2017), is perfectly compatible with a Kleinian view of the phantasy structures in which such hostile solidarity is articulated.

Similarly, the 'ontological insecurity' of the neoliberal period has surely been provoked by the creation of *real world insecurity* – the 'restructuring' of the labour market, the privatisation of social housing and gentrification, privatisation of health and education, fatal cuts to local services, erosion of the welfare system – in short, the destruction of the social safety-net of solidarity (see Garland 2000; Ramsay 2012). All this has made real life much more insecure for people in reality, and it is no surprise if this has activated persecutory anxieties and phantasies. But there are two ways in which psychology remains ineluctable for political sociology: first, political circumstance can only work through the psychological structures and tendencies that are already there to be activated, given our psychological nature, so that no political sociology that denies the significance of individual psychology could get going (see Bhaskar, 1998); and second, the question of the psychological features that made possible the passive acquiescence in the project of destroying social solidarity in the first place presses itself, a question which seems likely to lead us back to the subject of liberal modernity's defensiveness against civic despair.

Acknowledgements

I thank Henrique Carvalho and Anastasia Chamberlen for their comments, conversations,

encouragement and patience, and the anonymous reviewers for their comments. Alan Norrie provided detailed comments on an earlier draft, as well as many fruitful and supportive conversations, and Chloe Kennedy made helpful comments on an earlier draft. Many fine conversations with Matt Sinnicks and with Jaakko Nevasto have been invaluable, as has Alec Hinshelwood's intellectual companionship.

References

- Adorno T (1974) *Minima Moralia*. London: New Left Books
- Ashworth A and Zedner L (2014) *Preventive Justice*. Oxford: OUP
- Bauman Z (2000) Social Issues of Law and Order. *British Journal of Criminology* 40(2)
- Becker H (1963) *Outsiders*. New York: Free Press
- Bertram M (2006) The Actor-observer Asymmetry in Attribution: A (Surprising) Meta-analysis. *Psychological Bulletin* 132(6)
- Bhaskar R (1998) *The Possibility of Naturalism*. London: Routledge
- Bion W (1967) Attacks on Linking. in *Second Thoughts: Selected Papers on Psychoanalysis*. London: Karnac
- Braddock L (ms.) Understanding Projective Identification. unpublished
- Brandom R (ms.) Reason, Genealogy and the Hermeneutics of Magnanimity. unpublished
- Brown A (2003) From Individual to Social Defences in Psychosocial Criminology. *Theoretical Criminology*, 7(4)
- Brudner A (2009) *Punishment and Freedom*. Oxford: OUP
- Carvalho H (2017) *The Preventive Turn In Criminal Law*. Oxford: OUP
- Carvalho H and Chamberlen A (2017) Why Punishment Pleases: Punitive Feeling in a World of Hostile Solidarity. *Punishment and Society*. Online First
- Cheliotis L (2013) Neoliberal Capitalism and Middle-class Punitiveness. *Punishment and Society* 15(3)
- Cohen S (2001) *States of Denial*. Cambridge: Polity
- Delgado J (1969) *Physical Control of the Mind: Towards a Psychocivilised Society*. Harper and Row
- Duff A (1986) *Trials and Punishments*. Cambridge: CUP
- Duff A (2001) *Punishment, Communication and Community*. Oxford: OUP
- Durkheim E (1893/1997) *The Division of Labour in Society*. Basic Books
- Ethell L (2010) *Narrative Identity and Personal Responsibility*. Plymouth: Lexington
- Farnham D (2008) A Hegelian Theory of Retribution. *Journal of Social Philosophy* 39(4)
- Feinberg J (1965) The Expressive Function of Punishment. *The Monist* 49(3) 397–423
- Foucault M (1977) *Discipline and Punish*. Pantheon

- Freud S (1953) *The Interpretation of Dreams*. Hogarth Press
- Gadd D and Jefferson T (2007) *Psychosocial Criminology*. London: Sage
- Gardner S (1993) *Irrationality and the Philosophy of Psychoanalysis*. Cambridge: CUP
- Garland D (1990) *Punishment and Modern Society*. Oxford: OUP
- Garland D (2001) *The Culture of Control*. Oxford: OUP
- Gert et al (2004) Hampton on the Expressive Power of Punishment. *Journal of Social Philosophy* 35(1)
- Giddens A (1991) *Modernity and Self-Identity*. Stanford, CA.: Stanford University Press
- Gilligan C (1982) *In A Different Voice*. Cambridge, Mass.: Harvard
- Hampton J (1992) An Expressive Theory of Punishment. in Cragg (ed.), *Retributivism and its Critics*. Stuttgart: Steiner
- Harcourt E (2017) Containment and Rational Health. *European Journal of Philosophy* DOI: 10.1111/ejop.12316
- Hegel G (1821/1942) *The Philosophy of Right*. Oxford: Clarendon
- Hinshelwood B (1995) Changing prisons: the unconscious dimension. In Cox and Cordess (eds) *Forensic Psychotherapy: Crime, Psychodynamics and the Offender Patient*. London: Jessica Kingsley
- Hobbes T (1651/1991) *Leviathon*. Cambridge: CUP
- Hollway and Jefferson (2000) *Doing Qualitative Research Differently*. London: Sage
- Hudson B (1987) *Justice Through Punishment*. Basingstoke: MacMillan
- Jost J et al (2004) A Decade of System Justification Theory: Accumulated Evidence of Conscious and Unconscious Bolstering of the Status Quo. *Political Psychology*, 25(6)
- Kant I (1929) *Critique of pure reason* (trans: Smith, Kemp). London: MacMillan.
- Kant I (1993) *Critique of practical reason* (trans: Beck, White). New Jersey: Prentice-Hall
- Kant I (1997) *Groundwork of the Metaphysics of Morals*. (trans. McGregor). Cambridge: CUP
- Klein M (1934/1975) On Criminality. in *Love, Guilt and Reparation, and Other Works 1921-1945*. London: Hogarth
- Klein M (1935/1975) A Contribution to the Psychogenesis of Manic-depressive States. in *Love, Guilt and Reparation*,
- Klein M (1940/1975) Mourning and Its Relation to Manic-depressive states', in *Love, Guilt and Reparation*.
- Klein M (1946/1975) Notes on Some Schizoid Mechanisms. in *Envy and Gratitude, and Other Works 1946-1963*. London: Hogarth
- Klein M (1948/1975) On the Theory of Anxiety and Guilt. In *Envy and Gratitude*.
- Klein M (1957/1975) Envy and Gratitude. In *Envy and Gratitude*.
- Lacey N and Pickard H (2015) To Blame or to Forgive?. *Oxford Journal of Legal Studies* 35(4)
- Lear J (1998) Restlessness, Phantasy and the Concept of Mind. in *Open Minded: Working Out the*

Logic of the Soul. Cambridge, Mass.: Harvard

Lear J (2017) *Wisdom Won From Illness*. Cambridge, Mass.: Harvard

Lee A (2017) Defending a Communicative Theory of Punishment. *Oxford Journal of Legal Studies* 37(1)

Lemert E (1951) *Social Pathology*. New York: McGraw-Hill

Lerner M and Simmons C (1966) Observer's Reaction to the "Innocent Victim": Compassion or Rejection? *Journal of Personality and Social Psychology*, 4(2)

Maruna S (2001) *Making Good: How Ex-convicts Reform and Rebuild Their Lives*. American Psychological Association

Maruna et al (2004) Disowning Our Shadow. *Deviant Behaviour* 25(3)

Marx K (1844/2000) Towards a Critique of Hegel's *Philosophy of Right*: An Introduction. In McClellan, *Karl Marx: Selected Writings*. Oxford: OUP

Matravers M (2011) Duff on Hard Treatment. In Cruft, Kramer and Reiff (eds.), *Crime, Punishment and Responsibility*. Oxford: OUP

McCullough et al (2013) Cognitive Systems for Revenge and Forgiveness. *Behavioural and Brain Sciences* 36

Mead G H (1918/1964) The Psychology of Punitive Justice. in *Selected Writings*. University of Chicago Press

Merle J-C (2009) *German Idealism and the Concept of Punishment*. Cambridge: CUP

Meyer J (1971) Crime Deterrent Transponder System. *IEEE Transactions on Aerospace and Electronic Systems*, AES-7(1)

Mitchell S and Black M (2016) *Freud and Beyond: A History of Modern Psychoanalytic Thought*. New York: Basic Books

Money-Kyrle (1955) Psycho-analysis and Ethics. in Klein, Heimann and Money-Kyrle (eds.) *New Directions in Psycho-analysis*. London: Tavistock

Moore M (1996) *Placing Blame*. Oxford: OUP

Morris H (1976) *On Guilt and Innocence*. Berkeley: University of California Press

Murphy J (1979) *Retribution, Justice and Therapy*. Dordrecht: Reidel

Murphy J and Hampton J (1988) *Forgiveness and Mercy*.

Nietzsche F (1887/1998) Second Treatise. *On the Genealogy of Morality*. Hackett

Norrie A (1991) *Law, Ideology and Punishment*. Dordrecht: Kluwer

Norrie A (2000) *Punishment, Responsibility and Justice*. Oxford: OUP

Petersen et al (2010) Evolutionary Psychology and Criminal Justice: A Recalibrational Theory of Punishment and Reconciliation. in High-Olesen (ed.), *Human Morality and Sociality: Evolutionary and Comparative Perspectives*. Palgrave

Ramsay P (2012) *The Insecurity State*. Oxford: OUP

Rashdall H (1900) The Ethics of Forgiveness. *International Journal of Ethics* 10

- Robinson and Gadd (2016) Annihilation Anxiety and Crime. *Theoretical Criminology* 20:2
- Rodl S (ms.) Love and Justice. unpublished
- Schlosser J (2015) *Narratives and Discursive Discipline in Prison*. New York, NY: Routledge
- Segal H (1979) *Klein*. Brighton: Harvester
- Speziale-Bagliacca (2004) *Guilt*. Hove: Brunner-Routledge
- Spierenberg P (1984) *The Spectacle of Suffering*. Cambridge: CUP
- Spillius E (2011) *A New Dictionary of Kleinian Thought*. Abingdon: Routledge
- Stevens A (2012) 'I Am the Person Now I was Always Meant to Be': Identity Reconstruction and Narrative Framing in Therapeutic Community Prisons. *Criminology and Criminal Justice* 12(5)
- Strawson P (1962) Freedom and resentment. *Proceedings of the British Academy* 48: 1–25.
- Tetlock P (1994) Political Psychology or Politicised Psychology? *Political Psychology* 15(3)
- Valier C (2000) Looking Daggers: A Psychoanalytic Reading of the Scene of Punishment. *Punishment and Society*, 2(4)
- Vlastos G (1969/1981) The Individual As Object of Love. In *Platonic Studies*. Princeton, NJ.: Princeton University Press
- Von Hirsch A (1976) *Doing Justice: the Choice of Punishment*. Northwestern University Press
- Williams B (1995) Naturalism and Morality, in Altham and Harrison (eds.) *World, Mind and Ethics* (Cambridge: CUP)
- Winnicott D (1960/1965) Ego Distortion in Terms of True and False Self. In *The Maturation Process and the Facilitating Environment*. London: Hogarth
- Winnicott D (1954/1975) The Depressive Position in Normal Emotional Development. in *From Paediatrics to Psychoanalysis*. London: Hogarth
- Wollheim R (1984) *The Thread of Life*. Cambridge, Mass.: Harvard
- Wollheim R (1993) *The Mind and its Depths*. Cambridge, Mass.: Harvard
- Wootton B (1959) *Social Science and Social Pathology*. Allen and Unwin