Forthcoming titles:

Pierre Destree (*editor*)
ARISTOTLE ON AESTHETICS

David Dolby (*editor*)
RYLE ON MIND AND LANGUAGE

Christopher Pulman (*editor*)
HART ON RESPONSIBILITY

Bernhard Weiss (*editor*)
DUMMETT ON ANALYTICAL PHILOSOPHY

Philosophers in Depth
Series Standing Order ISBN 978–0–230–55411–5 Hardback
978–0–230–55412–2 Paperback
(*outside North America only*)

You can receive future titles in this series as they are published by placing a standing order. Please contact your bookseller or, in case of difficulty, write to us at the address below with your name and address, the title of the series and one of the ISBNs quoted above.

Customer Services Department, Macmillan Distribution Ltd, Houndmills, Basingstoke, Hampshire RG21 6XS, England

Kant on Emotion and Value

Edited by

Alix Cohen
University of Edinburgh, UK

palgrave
macmillan

11

The Heart as Locus of Moral Struggle in the *Religion*

Pablo Muchnik

I want to explore a usually neglected notion in Kant's account of moral fall and regeneration in *Religion*: the notion of 'heart' (*Herz*). This notion belongs to a constellation of concepts that Kant develops for the purposes of moral imputation and the attribution of responsibility. The other chief components of Kant's conceptual framework are 'propensity' (*Hang*), 'character' (*Charakter*), and 'disposition' (*Gesinnung*).[1] Although interpreters have tended to use these notions interchangeably, understanding their proper meaning, function, and scope in Kantian ethics is essential to preserve the consistency of the doctrine of radical evil. To make good on this claim, I will begin by discussing the contributions that the notion of 'heart' makes to Kant's account of the human moral condition, and argue that it is irreducible to the other components of his conceptual framework. This notion, we will see, is crucial to the success of Kant's anthropological argument in *Religion* I and invites us to reevaluate the role emotions play in our moral lives.

1

Kant introduces the term 'heart' first in a footnote, and then in the body of the text. Here are the relevant passages:

> Now, if we ask, 'What is the aesthetic constitution (*Beschaffenheit*), the temperament so to speak of virtue: is it courageous and hence joyous, or weighed down by fear and dejected?' an answer is hardly necessary. The latter, slavish frame of mind (*Gemütsstimmung*) can never be found without a hidden hatred of the law, whereas a heart (*Herz*) joyous in the compliance with duty (not just complacency in the recognition (*Anerkennung*) of it) is the sign of genuineness in virtuous

> disposition, even where piety is concerned, which does not consist in the self-torment of a remorseful sinner..., but in the firm resolve to improve in the future. This resolve, encouraged by good progress, must needs effect a joyous frame of mind (*Gemütsstimmung*), without which one is never certain of having gained also a love for the good, *i.e.*, of having incorporated the good into one's maxim. (R 6:23n)

And,

> Here, however, we are talking of a propensity to genuine evil, i.e., moral evil, which, since it is only possible as the determination of a free power of choice and this power for its part can be judged (*beurteilt*) good or evil only on the basis of its maxims, must reside in the subjective ground of the possibility of the deviation of the maxims from the moral law. And if it is legitimate to assume that this propensity belongs to the human being universally (and hence to the character of the species), the propensity will be called a natural propensity of the human being to evil. – We can further add that the will's (*Willkür*) capacity or incapacity arising from this natural propensity to adopt or not to adopt the moral law in its maxims can be called *the good or the evil heart* (*Herz*). (R 6:29)

The context of these passages is very different: in the footnote, Kant's focus is on the individual's pursuit of virtue; in the body of the text, his attention is directed to the propensity to evil which, he believes, can be attributed to the whole species. The alleged universality of the propensity motivates Kant's claim that 'the human being is evil by nature'. But this claim poses a problem for the first passage: if the propensity is indeed universal, how can it leave room for our acquiring good maxims? How can we be soiled by 'the foul stain of our species' (R 6:38) and yet have a good will?

These questions point at an implicit tension in Kant's view.[2] If Kant were to embrace a robust conception of the propensity to evil, it would seem then that the character of the species must determine the character of its individual members.[3] In that case, however, species affiliation would put an end to our freedom and the complaint of the Apostle would be true without exception: 'There is no distinction here, they are all under sin – there is none righteous (in the spirit of the law), no, not one' (R 6:39).[4] If, instead, Kant were to embrace a robust conception of individual freedom (a conception without which the pursuit of virtue and moral conversion are futile), it is then the universality of

the propensity that seems hopelessly compromised: since each human being will be free to adopt her moral character, some of us will be evil, but all will not be.[5] The first alternative is unacceptable. Sacrificing freedom would not only exculpate agents from having adopted their own character, but also undermine the basis of their mutual respect – an attitude grounded in the fact that 'rational nature exists as an end in itself' (G 4:428) and that our end-setting capacity is not susceptible to external constraints (MS 6:381). The second alternative, however, is no less problematic. Given our tendency to be deceived 'never more easily than in what promotes a good opinion of oneself' (R 6:68), and given the increasing lack of opportunity for wrongdoing that results from the consolidation of the coercive power of the state (PP 8:376n.), without full awareness of the propensity to evil nothing could awake us from our 'moral slumber'. We would be prone to 'fancy that [we] deserve not to feel guilty of such transgressions as [we] see others burdened with, without however inquiring whether the credit goes perhaps to good luck' (R 6:38).

The tension I am describing puts Kant in a bind: if he weakens the corrupting effects of the propensity, self-righteousness follows; if he budges on freedom, morality crumbles. What is remarkable in *Religion* is that such a dilemma never arises. The second passage quoted above makes this clear: despite its universality, the 'natural propensity to evil' does not preclude the individual's 'capacity or incapacity... to adopt or not to adopt the moral law in [her] maxims' (R 6:29). Kant makes the same point a few pages earlier: when we say that the human being is good or evil by nature, 'we are entitled to understand not individuals (for otherwise one human being could be assumed to be good, and another evil, by nature), but the whole species' (R 6:25). Kant evidently thought he could have it both ways: the universality of the propensity is in his mind perfectly compatible with some agents being good and others evil – instead of curtailing our autonomy, the propensity to evil manages somehow to cohabit with it. How is this possible? How did Kant eschew the latent contradiction between these claims?

There is no need to beat around the bush for an answer: Kant concludes the passage we are considering by explaining that it is in terms of the way an agent exercises her capacity to incorporate the moral law in her maxims that she can be said to have a good or an evil heart (*'das gute oder böse Herz gennant werde'*). So conceived, the individual's heart follows from, and expresses, her choice of character, but is not to be confused with that choice. Since the heart is for Kant an emotional expression of the principles an agent has adopted to govern her conduct (R 6:23n.), it

introduces a sensible content that illustrates those principles but is not identical with them. Seen this way, 'heart' becomes crucial to understanding how Kant's account could hold, without 'wobbling' or 'inconsistency', seemingly contradictory theses: namely, that the propensity to evil is universal (since it affects the heart of all human beings – the virtuous, the morally weak, and the vicious alike), and that each individual (understood now as an autonomous agent) is free to adopt her own moral character, irrespective of the character we may ascribe to the species. There is no contradiction between these claims, for 'heart' points at empirical regularities in the phenomenal manifestations of freedom, while 'character' refers to a noumenal act, which can be said to underlie and ground those regularities, but which is not part of the empirical/causal nexus.[6]

This distinction is important. If Kant were to capitulate on freedom, moral imputation would lose its point: 'The human being', Kant tells us, 'must make or have made himself into whatever he is or should become in a moral sense, good or evil. These two characters must be an effect of his free power of choice, for otherwise they could not be imputed to him and, consequently, he could be neither morally good nor evil' (R 6:44). Yet, and this is crucial, since that choice hints at an act (*Tat*) of transcendental freedom through which the individual agent is said to have established her moral character, independently of temporal/causal determinations, there is no contradiction between that act and assuming the universality of the propensity. For, Kant argues, 'the term "deed" (*Tat*) can in general apply just as well to the use of freedom through which the supreme maxim (either in favor of, or against, the law) is adopted in the power of choice, as to the use by which the actions themselves (materially considered, *i.e.*, as regards the objects of the power of choice) are performed in accordance with that maxim' (R 6:31). The two meanings of 'deed' are of course related – the first is supposed to ground the second. Yet, Kant makes it clear, talk of the propensity has phenomenal freedom in mind, since it refers to a feature that anthropological research gives us 'no cause for exempting anyone from [having]' (R 6:25), not to a (putative) noumenal act of moral self-constitution.[7]

Kant's conceptual bifurcation (*i.e.*, the two meanings of 'deed' and the two senses of freedom that correspond to them) keeps at bay the latent contradiction between the universality of the propensity and the freedom of the individual. Once we sufficiently distinguish between Kant's units of moral analysis (the species and the individual) and the kind of discourse that properly belongs to them (anthropology and morality), the tension in his doctrine of radical evil disappears. To the

extent that anthropology deals with the phenomenal manifestations of freedom ('deed' in the second sense) and morality with its noumenal underpinnings ('deed' in the first sense), Kant has sufficient conceptual room to hold that some individuals may be good and others evil, while the species, considered now empirically and through the lens of anthropology, gives us no reason to absolve it from having the propensity. Each claim is true within its own domain – a move that should remind us of the strategy Kant used in the first *Critique* to solve the third antinomy, but which he now applies to two different aspects of freedom itself.

This new application, however, brings its own set of problems. For, according to the logic of imputation, if the propensity is to qualify as genuinely 'evil', it must itself be the result of an act of transcendental freedom. This assumption puts enormous pressures on the empirical discourse of anthropology, forcing it to move in the direction of morality and incorporate elements of Kant's *a priori* conceptual apparatus. As a result of these pressures, I have argued elsewhere, Kant is led to develop a new brand of discourse, *moral anthropology*, which is neither completely empirical nor completely *a priori*, but combines elements of both.[8] My goal in this chapter is to discuss the role of the heart in that hybrid discourse.

2

According to the reading I am proposing, the notion of heart works as a *tertium quid*, a third term, irreducible to the character of the individual, and yet, on account of its sensible/emotional dimension, offering empirical data that warrant anthropological generalization. The versatility of this notion allows Kant to claim that each agent is fully responsible for the character she has, while linking the individual and the species in ethically meaningful ways. To fulfill the first function, Kant must distinguish 'heart' from 'character' and '*Gesinnung*', notions that also pertain to the sphere of individual morality, but which, unlike 'heart', refer to self-imposed practical principles with nothing empirical about them. To fulfill the second function, Kant must show how the heart, though uniquely personal, contains anthropological implications that allow us to discover patterns of emotion and conduct belonging to humanity as a whole, not just to the individual.

Kant tackles the second question first. Immediately after connecting the heart with the propensity to evil, he proceeds to develop a typology of immoral outlooks designed to straddle both units of moral analysis. This typology is based on the premise that, if we consider 'the play of

the freedom of the will *in the large*, [we] can discover within it a regular course' (I 8:17). This is a major methodological assumption on Kant's part, shared by his writings on history and much of the *Anthropology*, namely, that the spectator can find significant regularities in what, from the agent's point of view, are contingent acts of choice. Interpreted thus, each type of heart expresses a distinctive moral outlook in the person who possesses it, and yet suggests general patterns of emotion and behavior for the anthropologist to tease out:

> First, it is the general weakness of the human heart in complying with the adopted maxims, or the *frailty* of human nature; second, the propensity to adulterate moral incentives with immoral ones (even when it is done with good intention (*Absicht*), and under maxims of the good), *i.e.*, *impurity*; third, the propensity to adopt evil maxims, *i.e.*, the *depravity* of human nature, or of the human heart. (R 6:29)

Agents are frail when they lack moral strength and have maxims which, though good in intention, are overpowered by strong immoral inclinations; the impure act according to duty, but need sensible incentives to do so; finally, depraved are those who make of the reversal of the order of priority, first insinuated by the impure, a matter of principle, corrupting thereby the very basis of moral judgment, even when their actions remain legal (R 6:29–30). Despite their obvious differences (at the level of feeling, moral self-conception, and conduct), Kant detects a common volitional pattern at their basis: all these agents have made 'self-love and their inclinations the condition of compliance with the moral law – whereas it is the latter that, as the *supreme condition* of the satisfaction of the former, should have been incorporated into the universal maxim of the power of choice as the sole incentive' (R 6:36). Although each agent is evil in her own way, the maxims of evil people share a common refusal to limit the pursuit of inclination to the demands of morality. Even if human beings neglect the moral law in countless ways, each act of negligence is for Kant a variation on a single noumenal theme, the theme of evil willing.[9]

The most troubling aspect of this volitional orientation is that it does not necessarily lead to observable wrongdoing:

> So far as the agreement of actions with the law goes, there is no difference (or at least there ought to be none) between a human being of good morals... and a morally good human being..., except that the actions of the former do not always have, perhaps never have, the

> law as their sole and supreme incentive, whereas those of the latter *always* do. (R 6:30)

So,

> if this dishonesty is not to be called malice, it nonetheless deserves at least the name of unworthiness. It rests on the radical evil of human nature which (inasmuch as it puts out of tune the moral ability to judge what we think of a human being and renders any imputability entirely uncertain, whether internal or external) constitutes the foul stain of our species – and so long as we do not remove it, it hinders the germ of good from developing as it otherwise would. (R 6:38)

Kant's conclusion is not surprising: since he believes that the propensity to evil hides behind a mask of good conduct, it is easy for him to infer that there must be many more evil people in our midst than we had initially thought. So many, that at times it seems futile to provide 'a formal proof that there must be such a corrupt propensity rooted in the human being' (R 6:32). Given 'the multitude of woeful examples that the experience of human deeds parades before us' (*ibid.*), and the fact that the appearances of virtue (*Tugendschein*) do not disclose the true moral state of our character, the conclusion that the propensity to evil affects the whole species seems to naturally follow.

Yet, there is good reason to remain skeptical about Kant's reasoning here: even if we were willing to endorse his anthropological analysis so far, the evidence provided falls short of supporting Kant's claim about the universality of the propensity. In order for such a conclusion to follow, the claim that 'according to the cognition we have of the human being through experience..., we may presuppose evil as subjectively necessary in every human being, even the best of us' (R 6:32), cannot possibly limit itself to the depraved, the impure, and the frail. Such gerrymandering of humanity excludes precisely 'the best of us', and is hence insufficient to hold 'with no exception'. To avoid this problem, Kant would need to show that virtuous agents, no less than evil ones, have also traces of the propensity in their hearts. Only then could Kant's generalization be true in its own anthropological terms – not to mention the more serious philosophical problem of whether such an empirical claim could ever provide the type of *a priori* justification Kant's account seems to require.[10] But no matter how hard we look in the ensuing paragraphs, the missing anthropological evidence is not forthcoming.

I do not think, however, that the evidence is truly missing: it is buried instead in the passage with which we began our discussion – the footnote where Kant links the notion of heart with an individual's virtue. What makes the evidence so elusive is that the passage occurs at the early stages of *Religion*, at a point in which Kant has not yet advanced the thesis about evil's radicalism. Compounding the trouble (though perhaps as consequence of its position in the text) is that Kant does not refer here to the propensity by its proper name, but speaks elliptically of Hercules' 'subduing monsters' (R 6:23n). This ellipsis is not an isolated event. In the *Metaphysics of Morals*, for example, Kant assesses the strength of virtue by 'the magnitude of the obstacles that the human being furnishes through his inclinations', and identifies those obstacles with 'vices', that is, 'the brood of dispositions opposing the law, [which] are the monsters he has to fight' (MS 6:405). Similarly, at the opening of *Religion* II, where Kant resumes the discussion of virtue he started in the footnote, the struggle against vice is framed in terms of fighting an 'invisible enemy' and overcoming 'the malice (of the human heart) which secretly undermines the disposition with soul-corrupting principles' (R 6:57). Talk of 'enemy' and 'monsters', it is clear from the context, is just Kant's figurative way of getting at the propensity to evil. The lack of an explicit reference in the footnote should not sidetrack us.

3

Kant added the footnote to the second edition of *Religion* as a response to Schiller's accusation that there was no room for grace in his ethical theory.[11] Kant's rebuttal consisted in distinguishing between the concept of duty, 'which includes unconditional necessitation' and stands in 'direct contradiction' with gracefulness, and virtue, 'the firmly grounded disposition to fulfill one's duty strictly' and whose beneficent consequences 'allow the attendance of the graces' (6:23n). Far from inducing, as Schiller feared, 'the frame of mind (*Gemütsstimmung*) of a Carthusian', virtue manifests itself through a joyous and resolute heart (*Herz*). The virtuous agent, Kant argues, is neither despondent nor reproachful. On the contrary, virtue induces self-reliance and courage, and is sustained by a genuine 'love for the good' (*ibid.*). If, in strict sense, 'character (*Charakter*) signifies that property of the will by which the subject binds himself to definite practical principles he has prescribed to himself irrevocably by his own reason' (A 7:292),[12] Kant is using 'heart' here to designate the aesthetic dimension of that property, the affective correlate of the governing principle she has given to her faculty of desire. This is

why he associates the heart with 'temperament' (*Temperament*), a term Kant uses to refer to the emotional qualities (the characteristic feelings, inclinations and moods) that prompt, though not determine, agents to behave in certain ways (A 7:286–91).[13]

I take Kant's point to be that just as a virtuous character entails a distinctive way of thinking (*Denkungsart*) about morality, it also entails a distinctive way of feeling for the good – an emblematic 'frame of mind' (*Gemütsstimmung*).[14] In the 'heart', Kant locates the emotional dimension of an individual's moral life: the distinctive feelings, desires, inclinations, and moods that accompany and express her moral outlook. The same, of course, should be said with respect to an evil character – a point we touched upon in the prior section but whose further elaboration we must forgo, since our goal is to understand how the propensity finds its way into the heart of 'even the best of us'.

By connecting the heart with good character, Kant presumably intends to highlight that the acquisition of virtue is not simply an intellectual affair: it requires the mobilization of our emotions and must engage the whole human being, not just her reason.[15] Virtue demands a holistic conception of agency – a conception that supersedes (or at least softens) some of Kant's traditional dualisms. For, although the heart is an expression of the moral character that produces it, its relation with the principles an agent 'has prescribed to [herself] irrevocably by [her] own reason' (A 7:292) is not one of unilateral determination. As Kant sees it, the frame of mind with which an agent carries on her duties not only discloses, but also reinforces and helps perpetuate her moral commitment. A 'slavish frame of mind' is incompatible with virtue, for dejection not only gainsays the self-esteem that arises from being one's own master, but also undermines the chances of future moral improvement. According to this view, sensibility should no longer be construed as the other of reason, as an external power putting the will at a crossroads 'between its *a priori* principle, which is formal, and its *a posteriori* incentive, which is material' (G 4:400). If virtue is to have a chance, sensibility must somehow have been subsumed and transformed into reason's ally.[16]

This shift regarding the place of sensibility in Kantian ethics is the result of strictly anthropological considerations on Kant's part: without the cooperation of our emotions, long-term commitment to duty would exhaust the moral resources of a creature like us. While duty can – and will at times – demand great sacrifice on our part (think, e.g., of the friend-of-man of Groundwork I who, devoid of any inclination, fulfills his obligation by 'tear[ing] himself out of deadly insensibility' (G 4:398)), a life centered around the display of moral worth would be ultimately unsustainable. Given the kind of creatures we are, purity of principle is necessary but not sufficient to pursue virtue over the long run. '[T]he firmly grounded disposition to fulfill one's duty strictly' (R 6.23n) needs the support of a strong and joyous heart to accompany the work of reason, a moral aesthetics to sustain the agent in her 'love for the good' (R 6:23n).[17]

Talk of 'love' might seem out of place here – after all, Kant believes that '[l]ove is a matter of feeling, not of willing, and I cannot love because I will to, still less because I ought to (I cannot be constrained to love); so a duty to love is an absurdity' (MS 6:401). There is, nonetheless, room for a different, non-pathological type of love in Kantian ethics. As he explains regarding the duty of beneficence in the *Metaphysics of Morals*: 'If someone practices it often and succeeds in realizing his beneficent intention, he eventually comes actually to love the person he has helped' (MS 6:402). While love is indisputably a feeling, this type of 'love' is based on duty, not on inclination: the emotion does not precede but crowns the moral effort.[18] Even if a 'duty to love is an absurdity', a love that arises *because* of duty, in *response* to the moral law, is not: such emotion does not compromise the purity of our motivational structure – it supervenes upon it. Such byproduct is psychologically necessary and ethically beneficial, because a commitment to duty that required continuous and systematic emotional sacrifice would crash human beings under the tyrannical weight of pure reason, putting finite rational creatures, who cannot ignore their sensible interests, in constant self-discord.[19]

The 'love for the good' that Kant attributes to the virtuous agent in *Religion* follows the same pattern. The alignment of feeling and reason the virtuous manages to achieve is partly the result of her stable moral character, partly expression of the fact that creatures like us get pleasure 'in realizing [their] intention'. Our way of thinking and our way of feeling about the good reinforce each other here – hence the joy one finds in what one does. This is how Kant puts it:

> [I]t is one of the inescapable limitations of human beings and their practical faculty of reason ... to be concerned in every action with its result, seeking something in it that might serve them as an end and prove the purity of their intention –which result would indeed come last in practice (*nexu effectivo*) but first in representation and intention (*nexu finali*). Now, in this end human beings seek something that they can love, even though it is proposed to them through reason alone. (R 6:7n)

The immediate implication of this anthropological principle is that the centrality an agent chooses to give to duty in her character cannot be reduced to the purity of her principles alone, for it in part depends on her ability to become emotionally invested in the objects those principles direct her to pursue.[20] She must find something love-worthy in the commands of reason; otherwise, her adherence to them would be short-lived. Insofar as the moral project is not limited to a series of morally worthy actions, but shapes our entire life, it demands the alliance of our intellectual and affective powers – cognition of the good no less than emotional response to such cognition. Internal conflict might be inevitable on a sporadic basis, but continuous dissonance would jeopardize our chance of leading a virtuous life.

'Two things', Kant tells us in the *Metaphysics of Morals*, 'are required for inner freedom: being one's own master in a given case... and ruling oneself, that is, subduing one's affects and governing one's passions' (MS 6:407).[21] The state that results from this self-possession Kant calls 'moral apathy'. He describes it as a state in which 'feelings arising from sensible impressions lose their influence on moral feeling only because respect for the law is more powerful than all such feelings together' (MS 6:408). This *loss of influence*, however, is far from *indifference*: creatures like us, for whom 'reason solely by itself does not adequately determine [their] will' (G 4:412–13), must be always prepared to face a rift between their subjective and objective reasons.

The true accomplishment of the virtuous agent is that such a rift seldom occurs. Potentially disruptive feelings lose their grip over her, not because she stops being affected by them, but because she has brought them under the spell of pure practical reason. Thus, far from dispensing with emotions (as the deontological caricature would suggest), Kantian virtue puts sensibility to work at the behest of reason: without the arousal of a genuine 'love for the good', the agent would soon become a slave of her duty; without the feeling of respect, she would never learn to love the good.[22] Moral apathy is the result of having successfully bridged the gap between feeling and reason – a bridge by means of which the agent defers, but cannot eradicate, the occasional need of heroic self-sacrifice.[23]

As I read the footnote in *Religion*, it is in the individual's heart where this mediation takes place. To the extent that emotions (the complex manifold of feelings, desires, inclinations, affects, and passions) have an unmistakable sensible component, Kant considers the heart to be an independent moral faculty, neither fully sensible nor fully rational. Not fully sensible, because the emotions an agent harbors have been partially released from the causal nexus and transformed by her practical reason;

nor fully rational, because such transformation does not go 'all the way down', but preserves the irreducibly physical quality feelings have in Kantian ethics. Conversant with both aspects of our humanity, the heart is a link between freedom and nature and offers a path for pure practical reason to become embodied.

Stretching the bounds of orthodox interpretation, Kant's notion of 'heart' makes the heterodox assumption of a fluidity between reason and sensibility, introducing a kind of moral schematism that links the abstract principles ruling an agent's character with the rich emotions that give content to her moral life.[24] In the virtuous heart, emotions are prepared for being determined by the *a priori* concept of duty – a process that adapts the rules of pure reason to the sensible conditions of our moral experience. Just as schemata in Kant's theoretical philosophy are neither concepts nor intuitions, but 'a third thing, which must stand in homogeneity with the categor[ies] on the one hand and the appearance[s] on the other, [in order to] make the application of the former to the latter' (KrV A138/B177), in practical philosophy the heart provides a common ground for sensibility and reason to cooperate and enhance each other. Adapting a slogan from Kant's theoretical philosophy, one might say that, thanks to the heart, pure practical reason is not empty and feelings are not blind.

4

There still remains an important loose end: Is Kant justified in claiming that, 'according to the cognition of the human being through experience' (R 6:32), the propensity to evil must be considered universal? Does the heart of the virtuous, no less than the heart of the wicked, show evidence of the propensity within it?

The story of Hercules, I believe, contains Kant's answer. 'Hercules', Kant says earlier in the footnote, 'becomes *Musagetes* [leader of the Muses] only after subduing monsters, a labor at which those good sisters shrink back in fear and trembling' (R 6:23n; my emphasis). The Muses find the labor horrifying, the story implies, because the *warring* Hercules is for Kant a symbol of the frightful power of pure practical reason, which 'issues its precepts unremittingly, without thereby promising anything to the inclinations' (G 4:405). This two-stage developmental model is important, for it indicates that the phenomenology of emotions at the level of virtue must be distinguished from the sensible affectation experienced in the feeling of duty, its necessary predecessor in the individual's moral development. Only a diachronical interpretation of the Hercules story

can do justice to the transformation an agent must undergo to become virtuous. So interpreted, the progress of sentiments Hercules experiences, from 'respect to the law' to 'love for the good', offers a sensible representation of a noumenal improvement in his character, a sign of a change that would otherwise remain ungraspable.

The twist Kant gives to the Hercules story hinges on the assumption that the struggle against 'monsters' first starts as a struggle with oneself. For, unlike most feelings, respect is self-wrought: it registers the effect the moral law has on us, not the way we are affected by external objects. The effect is complex: the law initially produces pain, Kant argues, because it thwarts inclinations and 'humiliates' self-love; yet, this feeling is superseded by self-approbation, as we realize that the law is self-legislated, and the pain becomes a means to highlight the dignity of our rational nature (KpV 5:73–5, 81). The self-cancellation of pain serves to foreground Kant's ideal model for a proper self-relation: it reminds us that our pathologically determinable self has a limited and subordinate role to play in our moral personality (KpV 5:74).

There is no place for gracefulness at this stage of an agent's moral development, for 'the majesty of the law (like the law on Sinai) instills awe... [and] rouses a feeling of sublimity of our own vocation that enraptures us more than any beauty' (R 6:23n). '*Sublime*', Kant explains in the third *Critique*, 'is what, by its resistance to the interest of the senses, we like directly' (KU 5:267) – like respect, it produces pleasure by means of displeasure. The victorious Hercules, however, experiences morality differently when he becomes *Musagetes*. Having submitted the 'interest of the senses' to rational discipline, Hercules' sensibility has lost much of its former resistance to the law, bringing his feelings to engage in a more harmonious relation with his reason. Reconciled to its marginal and inferior role in the self-conception of a virtuous character, the agent's self-love loses much of its impetus 'to make its claims primary and originally valid, just as [if] it constituted our entire self' (KpV 5:74). In the struggle for self-mastery, Hercules becomes morally refined and develops a new form of love, 'a *love* for the good'. This transformation endears him to the Muses and allows him to discover a new kind of beauty in morality in lieu of the old sublimity. The experience of humiliation the agent undergoes, essential to the first moment of respect, prepares human sensibility to forfeit its claim to hegemony in the volitional structure of the virtuous person. Self-love is no longer a threat: as the source and organizing principle of our desires, feelings and inclinations, it has now a positive moral role to play as an assistant of duty.

To make possible this transformation, we have argued, Kant resorts to the notion of 'heart' – a hybrid notion whose assumption is necessary to extricate sensibility from its traditionally receptive/passive role and invest it with the spontaneous traits of agency which usually belonged to reason. This line of thought represents an important departure from a strand of argument dominant in the *Groundwork*: 'inclinations themselves', Kant used to believe, 'are so far from having an absolute worth, so as to make one wish to have them, that it must instead be the universal wish of every rational being to be altogether free from them' (G 4:428). This extirpation, Kant now realizes, would be a great drawback for morality. For, as we learn in the course of *Religion*, the 'monsters' Hercules vanquishes 'must not be sought in the natural inclinations, which merely lack discipline and openly display themselves unconcealed to everyone's consciousness, but [are] rather as it were an invisible enemy, one who hides behind reason and [is] hence all the more dangerous' (R 6:57). The enemy of virtue does not lie, as Kant used to think,

> in the sensuous nature of the human being, and in the natural inclinations originating from it. For not only do these bear no direct relation to evil (they rather give the occasion for what the moral disposition can demonstrate in its power, for virtue): we also cannot presume ourselves responsible for their existence (we cannot because, as connatural (*anerschaften*) to us, natural inclinations do not have us for their author). (R 6:35)

As mere physical forces, inclinations represent an opponent 'whose weapons cannot [possibly] touch' virtue – they are sensible magnitudes incapable of determining the agent's freedom.[25] So construed, inclinations are 'external' to morality and leave reason untouched. At most, they can 'make more difficult the execution of the good maxims opposing them; whereas genuine evil consists in our will not to resist the inclinations when they invite transgression, and this disposition is the really true enemy' (R 6:58n).

For Kant, the will not to resist is not mere weakness on our part, but a symptom of a fundamental depravity (*Bösartigkeit*) 'by which we throw dust in our own eyes and which hinders the establishment in us of a genuine moral disposition' (R 6:38). The image of the victorious Hercules, therefore, is meant to give sensible representation of what it means to regain moral vision. The monsters he must face are more formidable than the Hydra or the Nemean Lion, for evil is 'as it were

an invisible enemy, one who hides behind reason and [is] hence all the more dangerous' (R 6:57). If it is true, as Kant holds, that 'in Adam we have all sinned' (R 6:42, quoting *Romans* 5:12), in the Hercules favored by the Muses we find an image of what it means to overcome the *will not to resist*, the self-incurred moral depravity that Kant identifies with the propensity to evil. In Kant's hands, Hercules' labors represent the slow and uphill battle the individual must fight to attain virtue, a process in which the progress of one's sentiments reflects the progress of one's reason (MS 6:408).

Just as the Scriptures, in the form of a historical narrative, use the image of the serpent to add 'a closer determination of the depravity of our species by projecting evil at the beginning of the world, not, however, within the human being, but in a spirit of an originally more sublime destiny' (R 6:43–4), the story of Hercules presents in the shape of external 'monsters' what is in truth a projection of our internal struggle against the propensity to evil. Through Hercules' victories, we learn to imagine what it means to 'stand up' from Adam's fall, and through the love of the Muses, the meaning of self-overcoming.

5

Read this way, the footnote provides the piece of anthropological evidence missing in the body of the text: even a Hercules, 'the best of us', must prevail over the propensity to evil to become virtuous. The diachronicity of the narrative, the before and after experience of 'subduing monsters', leaves a mark in the heart of the individual: it creates a memory against which to measure moral progress. The heart is important in this process, for it gives representation and sensible expression to the otherwise ineffable nature of one's moral principles. If virtue did not have this kind of 'aesthetic constitution', we would neither be able to determine the nature of our character, nor if our commitment to its principles is genuine or spurious. Dispositional features that did not leave affective traces within us would remain ineffable and concealed, below the threshold of conscience, for we 'cannot observe the maxims upon which [they rest], we cannot do so unproblematically even within ourselves' (R 6:20). Without a heart, the agent would lack a sense of her evolving moral identity.

The need of sensible signs is due to the fact that duty can only show its motivational primacy indirectly, through what an agent feels and does, '*i.e.*, [through] the conduct of [her] life as it steadily improves, and from that [she] has cause to infer, but only by way of conjecture, a fundamental improvement in [her] disposition' (R 6:68). The inference is conjectural, for it moves from a given effect (the distinctive 'frame of mind' (*Gemütsstimmung*) that accompanies her virtuous conduct) to its alleged cause (the self-imposed guiding principle that governs her character). This type of reasoning, however, does not amount to certainty, since there is no guarantee that other factors might not have been involved in the process and contributed to the result. Yet, no matter how conjectural, Kant believes that the inference is sufficient to impart confidence and joy in the heart of the virtuous agent, who can 'reasonably hope that in this life he will no longer forsake his present course but will rather press in it with ever greater courage, since his advances, provided that their principle is good, will always increase his strength for future ones' (R 6:68).

Avoiding 'the sweetness or the anxiety of enthusiasm (*Schärmerei*)' (*ibid.*), the self-affirming power of virtue allows the agent to remain persuaded of her strength, but never overconfident about her powers to face the internal enemy. For she knows that

> certainty with respect to [her character] is neither possible to the human being, nor, so far as we can see, morally beneficial. For (be it well noted) we cannot base this confidence upon an immediate consciousness of the immutability of disposition, since we cannot see through to the latter but must at best infer it from the consequences that it has on the conduct of our life. (R 6:71)

Less certain than the self-righteous, but more confident than the forlorn, the virtuous agent finds in her heart resources to believe in her capacity to 'improve in the future' (R 6:23n). Hers is a kind of practical faith without objective warrant, a subjective belief that is nonetheless rational for her to have.

This explains why Kant considers the scrutiny of one's heart the first command of all duties to oneself:

> This command is 'know (scrutinize, fathom (*ergründe*)) yourself,' not in terms of your natural perfection ... but rather in terms of your moral perfection in relation to your duty. That is, know your heart (Herz) – whether it is good or evil, whether the source of your actions is pure or impure, and what can be imputed to you as belonging originally to the substance of a human being or as derived (acquired or developed) and belonging to your moral *condition*. (MS 6:441)

The demand for self-knowledge, Kant recognizes, is very taxing. The difficulty is not limited, as we have seen so far, to the uncertainty with which agents get to know their moral character (*Charakter*). It is compounded by the inscrutability of our *Gesinnung* – a notion Kant defines as 'the first (erste) subjective ground of the adoption of [our] maxims' (R 6:25). Although Kant often uses 'character' and '*Gesinnung*' as synonymous (both notions, after all, designate self-imposed higher-order principles that govern the individual's selection of maxims of action), *Gesinnung* has also a technical sense that is important to understand the radicalism of evil.

'This evil is radical', Kant tells us, 'since it corrupts the ground of all maxims' (R 6:37). 'Radical' is thus a spatial metaphor to designate the locus of evil at deepest level of maxim-formation in the *Gesinnung*, not the intensity of the wrongs we do.[26] The process of extrication, as we have seen, is very painful, since it involves a sweeping transformation of our moral self-image, a fundamental change in our way of thinking and feeling about morality. Thus, Kant concludes the passage from *Metaphysics of Morals* we just quoted:

> Moral cognition of oneself, which seeks to penetrate into the depths (the abyss) of one's heart which are quite difficult to fathom, is the beginning of all human wisdom. For in the case of the human being, the ultimate wisdom, which consists in harmony of a human being's will with its final end, requires him first to remove the obstacle within (an evil actually present in him) and then to develop the original predisposition to a good will with him, which can never be lost. (Only the descent into the hell of self-cognition can pave the way to godliness). (MS 6:441)

The way to godliness requires a descent into hell, because it is only after having accepted full responsibility for allowing the seeds of evil to take root in our *Gesinnung* that we can hope to cultivate the original predisposition to the good within us. Only after having dislodged the propensity to evil can humans begin to cultivate a good character (*Charakter*). The order of concepts is quite clear: the agent must 'first remove the obstacle within', that is, restore her moral ability to judge, and only after such restitution has taken place, can she develop 'the original predisposition to a good will with him, which can never be lost'.

This order indicates that '*Gesinnung*' and 'character' operate – as it were – as stacked concepts in the Kantian account of moral regeneration: while '*Gesinnung*' designates the act of self-constitution of our moral

agency, that is, the inaugural choice that 'applies to the entire use of freedom universally' (R 6:25) and is said to be responsible for the *radical* nature of the propensity, 'the grounding of character is like a kind of rebirth, a certain solemnity of making a vow to oneself; which makes the resolution and the moment when this transformation took place unforgettable to him, like the beginning of a new epoch' (A 7:294). Few human beings, Kant believes, 'have attempted this revolution before the age of thirty, and fewer still who have firmly established it before they are forty' (*ibid.*). That is, while the *Gesinnung* points to an act of freedom whose consequences we recursively discover by reflecting on how we have become the moral subjects we now are, the choice of character is forward-looking and expresses who we want to be as moral agents. Only after facing 'monsters' and dislodging the propensity *not to resist* the pull of inclinations can we start treading upon the slow path toward virtue.

How this revolution in the disposition is possible, 'surpasses every concept of our' (R 6:45); *that* it can take place is a *sine qua non* for moral imputation and the attribution of responsibility.

6

Much more needs to be said about the intricacies of moral conversion. But my goal has not been to discuss this vexing problem, but to call attention to the 'heart', a notion usually neglected by interpreters of Kant's ethical theory. My claim has been that, only if we avoid confusing 'heart' with the notions of 'character', '*Gesinnung*', and 'propensity', with which it shares a certain family resemblance, can we begin to make sense of Kant's doctrine of radical evil. The 'heart' dissolves the tension between Kant's apparently contradictory commitments to the universality of the propensity and the freedom of the individual. It does so by mediating, at the level of individual morality, between the *a priori* principles that constitute an agent's character and their phenomenal expression in typical moral emotions – a mediation that in turn makes those emotions susceptible to anthropological generalization. If this reading is correct, the 'heart' is the linchpin of Kant's moral anthropology – it is essential to understand his views in *Religion* about radical evil and moral regeneration, for the heart is the epicenter of an agent's moral struggle. Very much like imagination and the schematism associated with it, the heart makes the old Kantian dualisms look pretty obsolete – if not in the 'grounding' of morality, at least in its application to finite rational creatures like us.

Notes

This chapter has been written under the auspices of a Faculty Advancement Grant from Emerson College. I am thankful for the institutional support. An earlier version benefitted from the insightful comments of Lawrence Pasternack, Gordon Michalson, and Lauren Barthold. To them goes my gratitude.

1. I do not consider 'predispositions' (*Anlagen*) part of this group: although their activation and deployment is an act of freedom, predispositions 'are *original*, for they belong to the possibility of human nature' (R 6:28). They are 'constituent parts required for it' and hence 'belong with necessity to the possibility of this being' (*ibid.*). The notions we will be discussing, on the other hand, are contingent and acquired (brought about by us), and thus call for a separate grouping.
2. See Michalson (1990), 8–10, and Shulte (1991), 78–88. Michalson argues that this tension is a result of Kant's straddling two different traditions, orthodox Christianity (with its commitment to the doctrine of original sin) and Enlightenment values (with their commitment to autonomy). This set of influences introduces a series of inevitable 'wobbles' into Kant's account. Schulte has a harsher view: he argues that the universality of the propensity is incompatible with human freedom. The problem for him is not so much conceptual instability, but outright inconsistency.
3. O'Connor (1985), 288–309, and Frierson (2003), 95–135.
4. As I see it, this sweeping condemnation of human beings is the logical consequence of conflating the morality of the species and the morality of the individual. Unfortunately, such conflation is widespread among contemporary interpreters of Kant. For example, Henry Allison considers the '*Gesinnung*' (which he uses interchangeably with 'character') as synonymous with the 'propensity to evil'. He says: '[T]he distinctive features of the Kantian conception of *Gesinnung* are that it is acquired, although not in time, and that it consists in the fundamental or controlling maxim, which determines the orientation of one's *Willkür* as a moral being. Given this, we can now see that this *Gesinnung* is precisely what Kant means by a moral propensity of the conflation of the notions of 'evil disposition' (*böse Gesinnung*) and 'propensity to evil' (*Hang zum Bösen*)' (Allison (1990), 153). For views similar to Allison's, see Wolterstorf (1991), 41–53, and Timmons (1994), 114–44.
5. Guyer (2005), 115–45. This view is the flipside of the above: the emphasis on individual autonomy seems to weaken Kant's anthropological indictment.
6. For the 'grounding thesis', see Eric Watkins (Watkins (2005), 326). How such thesis affects 'the problem of moral anthropology' is central to Frierson's concerns (Frierson (2010b), 83–110).
7. The act is 'putative', for knowledge of freedom remains beyond our ken. Anthropology, on the other hand, is for Kant a *Beobachtungslehre*, an observational doctrine that deals with the phenomenal effects of freedom (Louden (2000); Louden (2011)).
8. Muchnik (2009a), 131–5; Muchnik (2013), 467–71.
9. The task of identifying this single volitional form is what Allen Wood has called 'the maxim problem' (Wood (2009), 150).
10. This has been a topic of heated controversy in recent years. See Allison (1990), Wood (1990), and (2010), 594–613, Morgan (2005), 63–114, and Muchnik (2009b), 116–43.

11. Schiller (1993). It is not clear, however, that Schiller really meant his comments as criticism. See Guyer (1993), 354; and Beiser (2005), 171.
12. I resort to Kant's definition of 'character' in *Anthropology*, because in *Religion* he does not provide one.
13. Kant does not have a single word comprising affects, passions, feelings, inclinations, and desires –each of which has distinctive meanings (even if Kant sometimes ignores his own distinctions). In this chapter, following what is customary in the literature, I will use 'emotions' (and its variants) as shorthand to refer to this complex manifold of terms. See Sorensen (2002), 109–28; and Borges (2004), 140–58.
14. The translation of '*Gemütsstimmung*' as 'frame of mind' is, to my ear, too intellectualistic and does not capture the emotional/affective undertones that '*Stimmung*' has in German. 'Mood' is perhaps a more appropriate rendering.
15. See Guyer (1993); Sherman (1990), 149–70, and (1995), 369–77.
16. A similar transformation of the role of sensibility occurs in the *Metaphysics of Morals*, where Kant speaks of '*aesthetische Vorbegriffe*' (translated as 'what is presupposed on the part of feeling') 'by the mind's (*Gemüts*) receptivity to concepts of duty as such' (MS 6:399). See Guyer (2010), 130–51.
17. So construed, the heart brings additional support to an 'affectivist' interpretation of Kantian ethics (see McCarty (1993), 421–35). Such interpretation challenges the standard 'intellectualist' reading of the feeling of respect, according to which the sheer intellectual recognition of the moral law is in itself sufficient to motivate moral conduct. What the heart adds to the affectivist line of thought is twofold: (a) it shifts its attention from isolated action to the whole of an agent's life, and (b) it extends the repertoire of morally relevant feelings, from 'respect' to the more unusual feelings of 'love', 'courage', 'resolution' and 'joy'.
18. The command from Scripture 'to love our neighbor, even our enemy' (G 4:399) as we love ourselves would be incomprehensible otherwise. The love in question is the result of self-mastery, a sign of having overcome the unloving feelings our neighbor naturally produces.
19. To the extent that the moral emotions here do not precede, but follow from the choice of character (a choice that must be conceived to have taken place independently from phenomenal determination), feelings support, but do not change the type of reason an agent is prone to have as a result of that choice. This type of cooperation between our phenomenal and noumenal character, therefore, preserves the direction of determination of Kant's 'grounding thesis' (Watkins (2005)). It should not be confused with what Frierson calls 'the problem of moral anthropology', namely, the use of empirical causes (moral education, the use of social incentives, *etc.*) to modify the kind of reasons an agent is liable to have as a *deliberator* (Frierson (2010b), 103–7). This curious alchemy of *causes* into *reasons* reverses the direction of noumenal determination and introduces a different set of problems than the ones that occupy us here.
20. There is another, far-reaching implication of this observation, which Kant uses to explain the emergence of the doctrine of the highest good out of the very core of morality (Muchnik (2009a), and Pasternack (2014)).
21. Despite their shared sensible underpinnings, affects (*Affekten*) and passions (*Leidenshaften*) are very different for Kant (see MS 6:406 and A 7:252ff). I leave the discussion of their difference for another time.

22. Kant insists that the love for the good be rationally mediated, for otherwise it rests on affects and falls into 'enthusiasm' (see MS 6:409).
23. See Engstrom (2002), 289–315.
24. See Muchnik (2009a), 139–42.
25. I borrow the idea of 'weapons' from Jeanine Grenberg (Grenberg (2012), 152–69). Grenberg persuasively shows how Kant's understanding of inclinations as physical forces, which often surfaces in the *Metaphysics of Morals*, must in fact be interpreted in light of his discussion of 'vices', and is hence compatible with his view of evil and freedom in *Religion*.
26. Henry Allison and Allen Wood make similar points regarding the meaning of 'radical' (Allison (1990), 153; Wood (1999), 284).

12

Kant and the Feeling of Sublimity

Michelle Grier

As with a number of theorists in the modern period, Kant takes the sublime to center on a unique and complex aesthetic experience, one in which our delight in an object is complicated by attendant, and contrary, feelings of pain, and repulsion. Although Kant does sometimes suggest that some works of fine art can *depict* the sublime (a Romantic painting of a sublime landscape), or present ideas *in a manner* that is sublime (especially the arts of speech), his considered view is that the sublime is best sought in 'crude nature', such as the 'wide ocean, enraged by storms', where our feelings of sublimity are occasioned immediately.[1] Thus, confining ourselves to Kant's claims about the sublime in nature, we might say that the sublime is loosely associated with those aesthetic cases in which we feel ourselves overwhelmed or awed by natural phenomena that present as magnificent or mighty: 'it is rather in its chaos' that nature most arouses our ideas of the sublime, or in its 'wildest and most unruly disarray and devastation, provided it displays 'magnitude and might' (KU 5:246–7). The formlessness in objects *often* associated with the sublime in nature highlights what is for Kant one significant difference between the sublime and the beautiful.[2] For whereas natural beauty allows us to present nature as a system 'in terms of laws whose principle we do not find anywhere in our understanding', experiences of the sublime in nature yield no particular 'objective principle' to which the forms of nature conform (KU 5:246). Whereas natural *beauty* leads us (rightly) to judge appearances as belonging not merely to nature as a mechanism without purpose, but also to belong to nature considered by analogy with art (*i.e.*, as purposive for our judgment), the *sublime* leads us to judge appearances as, as it were, 'contrapurposive' for our aesthetic power of judgment (KU 5:246). Indeed, the experience of the sublime serves rather to catapult us into a recognition of a different (presumably