

# Nature, Corruption, and Freedom: Stoic Ethics in Kant's *Religion*

Melissa Merritt, University of New South Wales

## 1. Introduction

This paper is an invitation to take seriously Kant's debt to Stoic ethics and moral psychology. Scholars have given some attention to this connection, though often in the form of drive-by remarks, where, often for perfectly good reasons, the matter cannot be fully taken up.<sup>1</sup> When it is taken up, scholarly caution sometimes recommends a deflationary verdict, citing either uncertainty about Kant's access to Stoic texts, or systematic differences.<sup>2</sup> But there is no great mystery about Kant's access to Stoicism, which comes primarily through Seneca and Cicero, as it has for most philosophers since the Renaissance. And while there are genuine systematic differences — Kant, for example, is not a eudaimonist<sup>3</sup> — I would argue that his moral psychology, in particular, has clear and profound Stoic roots. Of course, a full argument for the influence of Stoicism on Kant's ethics and moral psychology is a bigger project than can be undertaken here: my present aim is to make a case for Kant's debt to the Stoics on one topic, in one work.

Kant's account of "the radical evil in human nature" in Part 1 of *Religion within the Bounds of Reason Alone* is typically interpreted as a reworking of the Augustinian doctrine of original sin, in a manner intended to establish its compatibility with Enlightenment morality. But Kant

---

<sup>1</sup> E.g. in work primarily in ancient philosophy, such as Cooper (2004), or the relatively more developed remarks in Nussbaum (1997), and Annas (1993, *passim*); or in Kant scholarship outside of moral philosophy, such as Allison (2001, 344).

<sup>2</sup> Schneewind (1996) cites a range of systematic differences to suggest that, in effect, the question of the Stoic influence on Kant's ethics does not deserve serious study, despite observations on the connection in Annas (1993). He groundlessly asserts that Kant gets his Stoicism chiefly through Leibniz, when there in fact there is ample evidence that Kant engaged directly with ancient sources of Stoicism — as his education (see Kuehn 2001, 48-49), his teaching (consider, e.g., the prominence of Stoic ethics throughout the Vigilantius lecture notes, Ak 27: 479-732 *passim*), and quotations in his own writing (some of which we shall consider here) all attest. From a different angle, Sherman (1997) has done much to advance our understanding of Kant's debt to Roman Stoicism; however she assumes that Kant must be committed to a dualist psychology more in the style of Aristotle, thereby taking emotions to be non-rational (see also Sherman [2013]). I explain below (§4) why this cannot be assumed for Kant's view of passions (*Leidenschaften*); but I do not here attempt to challenge Sherman's broader assumption that Kant rejects the psychological monism of the Stoics. This would require a separate study focused on larger systematic issues, whereas here my particular focus here is on the significance of Stoic ethics for Kant's account of radical evil.

<sup>3</sup> Annas (1993) argues that effectively all ancient moral theories were eudaimonist, and provides nuanced discussion (see especially pp. 388-411) of the special way in which this true of Stoics, who take our happiness to consist in virtue, or the completion of our essentially rational nature. Annas (1993) is worth considering for its attention to the relation between Stoic and Kantian ethics on this and other points.

doesn't talk about Augustine explicitly there, and if he is rehabilitating the doctrine of original sin, the result is not obviously Augustinian.<sup>4</sup> At any rate, my aim is not to deny the relevance of the Augustinian background to Kant's account of radical evil, but rather to make a case that the neglected Stoic background might be at least as important.<sup>5</sup>

"Radical evil" is the idea that our moral condition is — by default and yet by our own deed — bad or corrupt; and that this corruption is the root (*radix*) of human badness in all its variety, ubiquity, and sheer ordinariness.<sup>6</sup> As I will argue, Kant's account of radical evil takes as its premise a version of the Stoic principle that nature gives us "uncorrupted starting points [ἀφορμὰς ... ἀδιαστρόφους]" (DL 7.89). So we must corrupt ourselves. But how, and why, do we do this if we are thus created? This was debated among Stoics in antiquity. And if Kant accepts the Stoic premise, it would be his problem, too. My overarching aim is to show how Kant's account of radical evil is the product of his engagement with this originally Stoic problem.<sup>7</sup> Let me provide an overview of the issues at stake.

---

<sup>4</sup> Kant explains his task as, in part, establishing the coherence of certain "fragments" of revealed religion with rational religion (Rel 6:12); and plausibly he takes the doctrine of original sin as the first of these "fragments", although he arguably only makes oblique reference to the doctrine (*peccatum originarium*, 6:31). Hence commentators debate whether Kantian radical evil is even a conception of original sin: compare, e.g., Wood (2014) and Guyer (2009). It is perhaps more widely accepted that Kantian radical evil is not, at any rate, an *Augustinian* conception of original sin: on this, see Pasternack (2017, 462), and for an example of the Augustinian reading, consider Mariña (1997). By my lights, Kantian radical evil bears a closer resemblance to the Pelagian view that sin is a matter of choice, and thus that we have not passively inherited the sin of Adam (see Rel 6:40) — the doctrine the Church deemed heretical when it endorsed the Augustinian doctrine instead. Kant also clearly rejects the Augustinian idea that merit-independent divine grace is needed if we are to be restored to the good from our fallen moral condition. We will return to some of these themes, but it lies outside of my scope here to consider them in relation to Augustine.

<sup>5</sup> Relatedly, one might ask: why focus on the Stoics, rather than Rousseau, who is also clearly on Kant's mind in his account of radical evil? Certainly my aim is not to deny the importance of Rousseau to this topic; however, Rousseau figures more prominently when Kant focuses on the *social* dimension of human corruption in *Religion 3*, which lies outside of my immediate field of concern here. (On that point, though, I would argue that Kant is thinking at least as much about Seneca as about Rousseau.) See Brooke (2001) on Rousseau's complex relation to both Augustinian and Stoic views about human corruption; there is certainly important work to be done on how Kant's grappling with this dual inheritance might be informed by Rousseau's prior grappling with it — but that lies outside of my scope here. There is, I contend, sufficient evidence of Kant's going to Stoic sources directly to merit our trying to understand his relation to them on its own terms.

<sup>6</sup> "Radical evil" is therefore *not* the idea of "extreme evil"; for discussion of this common misconception, see Loudon (2010). Since already the English "evil" has some connotation of *extreme* depravity, I will generally prefer to speak of "badness".

<sup>7</sup> It is widely recognised that Augustine draws much from Stoicism in his account of original sin (for varying accounts see, e.g.: Frede (2011), ch. 9; Brooke (2012), ch. 1; Byers (2013), esp. ch. 7), even though he ultimately rejects Stoic ethics as blasphemous. Thus while it might be instructive to compare what Augustine and Kant each accept and reject from Stoic ethics and moral psychology, that is also not my project here.

The outline of any Stoic answer to the question of how we corrupt ourselves begins by linking the acquisition of badness with the acquisition of reason. Alexander of Aphrodisias, though a hostile critic, correctly reports the Stoic view that since children are not yet rational, they can be neither good nor bad; and, he continues, when children “change to being rational [μεταβάλλοντας δ’ εἰς τὸ λογικὸν]” — and so are no longer children — they are “straightway bad, but without becoming so [εὐθὺς κακούς, ἀλλ’ οὐ γίγνεσθαι]” (AA 122.3-5/SVF 3.537). We are thrown into rationality and badness at once, in a single stroke. Given what the Stoics take to be involved in the acquisition of reason, the relevant details of which we will consider later, the upshot is that we bring badness upon ourselves, and do so inevitably from the first exercise of reason. Badness is some kind of self-induced human default.

With this constellation of claims in view, the reader of Kant’s *Religion* might be struck by some similarities. Kant’s own account of radical evil begins with an assertion that there is an “original predisposition to good in human nature” (6:26): as I will argue, this is his version of the Stoic premise that nature gives us uncorrupted starting points. And like the Stoics, Kant insists that we can only corrupt ourselves. Moreover, he takes this corruption to be coeval with our coming into the use of reason, which in this context means acquiring the resources of a genuinely practical, agential point of view. Thus Kant says that radical evil can be called an “*innate* guilt (*reatus*) [...] because it is detectable as early as the first manifestation of the exercise of freedom in the human being, and yet nonetheless must have originated from freedom and is therefore imputable” (6:38).

So we have a shared premise, that nature predisposes us to the good, and a shared problem, to understand how we corrupt ourselves given this predisposition. There is also something shared in the line of response, inasmuch as it takes the acquisition of badness to be coeval with the acquisition, or the coming online, of reason.<sup>8</sup> Alexander takes this coequality claim to mean that, by Stoic lights, reason itself is the source of the corruption. Thus the Stoics, he continues, identify the “the change to being rational” with the change to being “bad”, and as a result implicitly take vice to be natural, and virtue unnatural, for a rational being (AA 122.5-7) — an absurd result that conflicts with the premise that nature predisposes us to the good. But Alexander overlooks the distinction, on which both the Stoics and Kant rely, between acquiring

---

<sup>8</sup> The Stoics think that we acquire reason at a certain age (roughly fourteen), once we acquire “the stock of common notions naturally shared by all human beings” (Frede 1996, 11), whereas Kant thinks that we are endowed with a rational capacity but yet *come into* its use — presumably with the dawn of genuine thought, which involves apperceptive self-consciousness (see Anth 7:127). Although there are important differences between Stoic and Kantian conceptions of reason, the very idea that the Stoic takes reason to be acquired at some point, whereas the Kantian thinks of a capacity that comes online at some point, does not seem to me to be one of them.

the resources of rationality and developing rationality to a state of completion. The change to being rational is a matter of acquiring certain mental resources; it is not the acquisition of perfected reason, or virtue. By analogy, there should be no equivalence between the acquisition of rationality and the acquisition of vice; and the Stoics do not implicitly take vice to be natural to us. Kant seems to have something like this rejoinder in mind when he sets out to explicate the saying that “the human being is by nature evil” and immediately rules out its meaning that evil “may be inferred from the concept of his species”; for if evil were inferred from our essentially rational nature, “then the quality would be necessary” (Rel 6:32). Although the rejoinder doesn’t solve anything, it nevertheless reframes the explanatory problem, which is to make sense of how we are inevitably, but not necessarily, bad.

When Kant denies that radical evil is a “necessary” quality of the species, he evidently means that it is not bound up in our essentially rational nature. Reason itself does not make us bad. But it may be that we are bound to misuse reason from the beginning. This is what Kant appears to have in mind when, in the very same sentence where he denies that radical evil is a “necessary” quality of the species, he nevertheless maintains that it is “subjectively necessary in every human being” (Rel 6:32). While radical evil is not to be traced to our rational nature as such, it is to be traced to the acquisition of a practical point of view: by Kant’s lights, we inevitably choose in such a way that brings the corruption upon ourselves just as soon as we come into the use of our reason, even though this acquisition brings with it resources that allow, in principle, for choosing otherwise.

The Stoic background of Kant’s topic is indicated in the *Religion* in various ways. Most notably, Kant explicitly discusses Stoic ethics in a pair of passages that border his own account of radical evil on either end: the second paragraph of the preamble to *Religion* Part 1 (6:19-20), and the entirety of the preamble to *Religion* Part 2 (6:57-59). For ease of reference, I refer to these passages as the “opening” and “closing” frames around Kant’s account. We examine them in (§2). Why does Kant insert his account of radical evil in that frame? The answer that immediately suggests itself — namely, that the frame is Kant’s nod to the Stoic origins of his topic — is correct as far as it goes. But it turns out that there is a significant shift in the portrayal of Stoicism across those passages — a point that has not been considered by the relatively few commentators who have paid attention to Kant’s discussion of Stoicism in this context.<sup>9</sup> This shift, I will suggest, reflects an intramural Stoic debate about how to explain our corruption; the

---

<sup>9</sup> Such as Baxley (2010, 71-2), and more indirectly Grenberg (2005, 15-21).

key player in this debate was the Stoic Posidonius, who was something of a teacher to Cicero, and frequently cited as a Stoic authority by Seneca.<sup>10</sup>

Crucially, then, the frame bears traces of the Stoic history of grappling with the problem of explaining human corruption. This complication — for there are, effectively, two Stoicisms in play — gives us a more interesting angle on the interpretive question about why Kant inserts his account of radical evil within this frame. Kant’s message, in effect, is that the Stoics have somewhat misconceived the problem and, as a result, have been misguided in their pursuit of a solution. Kant’s clarification of the problem is to reject the canonical Stoic supposition that reason must somehow corrupt itself: *what* gets corrupted, in Kant’s view, is instead the original predisposition to the good. This move does not make the problem of any easier to solve: it rather heightens our appreciation of the difficulty. In §3, we will see that the shared premise that we are originally predisposed to the good concerns our affective relationship to ourselves, and our proper attachment or attraction to what promotes the completion of our nature. If radical evil is the corruption of this predisposition, the upshot is that we are inevitably attached to our own faults, and prone to be indifferent or hostile to what promotes the completion of our own essentially rational nature. How, again, do we become so perverted? And once we are, what resources could we possibly have to rehabilitate the original predisposition on our own initiative?

By Kant’s lights, the appropriate way of appreciating and addressing these questions requires a conception of free choice that is not fully available in his Stoic sources. My account of his response begins by examining how his premise about our original predisposition to the good both draws upon and departs from the Stoic version of this premise and its elaboration in the Stoic doctrine of *oikeiōsis* (§3); then I explain how Kant provides an account of radical evil as the inevitable corruption of this predisposition (§4). Kant’s claims about the incomprehensibility of radical evil (Rel 6:21, 25, 32, 43) have been widely discussed by commentators; somewhat less attention has been given to his companion claim that “the restoration of the original predisposition to good to its power” is just as incomprehensible (Rel 6:44-45). This is a surprising move: after all, Kant provides some kind of account of both. Each time Kant claims their incomprehensibility, he cites the same ground: namely, that both are the result of free choice. In §5, I explain how his reply to the Stoic handling of the problem turns on this point.

\*

---

<sup>10</sup> See, e.g., Seneca Ep 87.35, 88.21, 104.22, 108.38, 113.28, and particularly Ep 121, on *oikeiōsis*, and Ep 90, on human history, where the question of how vice infiltrated society is implicated — both issues of concern in this paper. Kidd provides a brief biography in Posidonius (1999).

That is an overview of what I aim to establish in this paper. My inquiry will raise questions about the sophistication of Kant’s understanding of Stoicism, which will be addressed later on.<sup>11</sup> At this point, some elaboration on Kant’s access to Stoic texts may be in order. Kant owned a complete edition of Seneca’s moral essays,<sup>12</sup> from which he quotes in *Religion* 1. Apparently the only straight philosophical work of Cicero that Kant owned was *De Officiis* (*On Appropriate Action*) — not in Latin, but in Christian Garve’s 1783 controversial German adaptation of the work.<sup>13</sup> He did not own either *Tusculan Disputations* or *De Finibus* (*On Moral Ends*), two works which will concern us here. But the catalogue of Kant’s relatively small personal library (Warda 1922) does not tell us much, really, about which philosophical works mattered to him, as we can see from the following example. In *Tusculan Disputations* 2.61, Cicero tells the story of Pompey Magnus visiting Posidonius, when he stopped in Rhodes returning to Rome from a military campaign. He arrived to find Posidonius suffering from an attack of gout. Wishing not to intrude, Pompey made to leave — but Posidonius would hear nothing of it, taking the situation as the occasioning subject matter for his lesson. Many of Kant’s readers know this story, without realising it, from the *Critique of Practical Reason*: “one may always laugh at the Stoic who in the most intense pains of gout cried out: ‘Pain, however you torment me I will still never admit that you are something evil (κακόν, *malum*)!’; nevertheless he was correct” — *correct*, Kant goes on to explain, because he did not confuse “the worth of his person” with “the worth of his condition” (5:60). Kant does not cite his source or name “the Stoic” in the second *Critique*, but he does when he repeats the story in a later essay, “Announcement of the Near Conclusion of a Treaty for Eternal Peace in Philosophy” (8:414, citing precisely *Tusculan Disputations* 2.61). From this pregnant example, we can surely grant that the absence of a given text from Kant’s personal library warrants no inference about its lack of significance for his formation as a philosopher.<sup>14</sup>

## 2. The Stoic frame

We begin by looking into Kant’s explicit discussion of Stoic ethics on either end of his account of radical evil. The most immediate observation to make is that the opening frame expresses a

---

<sup>11</sup> Mostly in §§2-3.

<sup>12</sup> Seneca (1762) contains all the moral essays and consolation writings, but not the letters to Lucilius.

<sup>13</sup> J.G. Hamann reported in a letter that Kant was working on a reply to “Garve’s Cicero” — a work that eventually became the *Groundwork*: on the historical context see Kuehn (2001, 278-87); for an effort to trace the impact of the Stoic Panaetius (via *De Officiis*) on the development of the *Groundwork*, consider Reich (1939); and for a justly sceptical assessment of Reich’s effort, see Wood (2006, 361-364). (Kant did own a 1657 Latin volume of Cicero’s speeches.)

<sup>14</sup> Similar evidence is provided by the absence of *any* work by Rousseau in his personal library, whom he yet acknowledged as having had a profound influence on the development of his practical philosophy (Ak 20:44).

kind of moral optimism that goes missing in the closing frame. Our questions have to do first with the philosophical commitments that underlie this optimism, and why it is not sustained in the closing frame.

## 2.1 The opening frame

The very first paragraph of *Religion 1* claims the apparently universal, certainly cross-cultural, agreement that the world and human existence began good, and then fell into decrepitude and evil. “All allow” this beginning in a Golden Age or Paradise, and subsequent fall into evil — and most suppose that everything is only going completely to hell from here (Rel 6:18). What I am calling the “opening frame” is the second paragh of *Religion 1*, which expresses the “opposite heroic opinion [...] that the world steadfastly (though hardly noticeably) forges ahead in the very opposite direction, namely from bad to better; that at least there is in the human being the predisposition to move in this direction” (6:19-20).<sup>15</sup> Kant points to Seneca as one of its original representatives, and claims it has gained ground among Enlightenment pedagogues, like Rousseau. This optimistic view rests on the presupposition that there is a “seed of goodness”, indeed an “ethical predisposition towards goodness” in us (6:20). The passage closes with a quotation from Seneca that will help us understand the basis of this optimism:

[W]hat ails us is curable, and if we're willing to be freed of our flaws, nature herself lends a hand, for we are *born to be upright*.

sanabilibus aegrotamus malis, nosque *in rectum genitos* natura, si sanari velimus, adiuvat.

(*Religion 6:20*, slightly misquoting Seneca, *De Ira 2.13.1*)<sup>16</sup>

In the background of Seneca's remark is the original Stoic premise that nature gives us uncorrupted starting points (DL 7.89), or as Seneca himself says elsewhere “nature does not predispose us to any fault [Nulli nos vitio natura conciliat]; she has begotten us whole [integros] and free” (Ep 94.56). Put positively, the Stoic premise says that nature guides us towards our proper end, which is “to live in agreement with nature, which is the same as to live virtuously

---

<sup>15</sup> Kant's description of the view as “heroic” nods to his central (and persistent) criticism of Stoic ethics, that its sage is supposed to have *transcended human nature*; see KpV (5:127), and cf. Seneca (Ep 124.23-4).

<sup>16</sup> I quote Kaster's translation in Seneca (2010), since the rendering of Seneca in CEWIK is too free; I follow with Kant's own slightly misquoted version of the Latin, which he does not translate into German. Seneca has (1) *emendari* where Kant has *sanari* (“if we are willing *to be freed of our flaws*” instead of “... *to be restored to health*”) and (2) *ipsaque ... natura* (“nature *herself* lends us a hand”). Kant's Latin edition of Seneca (1762) agrees with our text, so the departures are likely attributable to Kant's quoting from memory. (Rousseau also used the line as the epigraph to *Émile*, but as far as I can tell it was always quoted correctly there.) I am very grateful to Brad Inwood for helping me with this bit of Seneca, and particularly the Latin; any mistakes are of course entirely my own.

[κατ' ἀρετὴν]; for nature leads us towards virtue [ἄγει ... πρὸς ταύτην ἡμᾶς ἢ φύσις]" (DL 7.87). Seneca invokes the positive claim here. Kant's additional stress on the phrase *in rectum genitos* emphasises that *virtue* is the telic goal, while his slight misquotation somewhat de-emphasises that *nature* leads us towards it.<sup>17</sup> So while nature endows us with this goal and apparently induces us towards it, she does not herself bring us there. We are born to *make ourselves* upright or right-acting.

The passage conceives of our corruption as some kind of disease — but a *culpable* one that we bring upon ourselves. By Stoic lights, the sickness is expressed in ordinary emotions or *pathē*, which are themselves expressions of assent to false evaluative propositions. So, for example, Seneca explains the *pathos* of anger as the felt expression of assent to the complex proposition that one has been done wrong by another, and that one ought to be avenged (*De Ira* 2.1.4). We take this injury to be genuinely bad (when the Stoics say it is not) and as calling for revenge (when the Stoics say that it does not). But therein lies the silver lining on which the opening-frame optimism rests. If we corrupt ourselves, then we should be able to undo what we have wrought. We can unlearn our mistakes. Seneca's message is that we chiefly need to be *willing* to be freed of our flaws — and who wouldn't want that?

In *De Ira*, Seneca continues, appearing somewhat exhilarated on the matter: “Nor is the path to the virtues steep and rough, as some have thought: the approach lies on level ground. I don't come to give you nonsensical advice. The way to the best human life is easy” (*De Ira* 2.13.1-2). But is it so easy, really, to want to be freed of our faults? Seneca frequently takes the part of a rousing coach; his moral essays and letters are distinctively hortatory. The fact that he so often finds it appropriate to speak in this vein is itself some testament to his recognition that this business might not be so easy after all. We are liable *not* to be willing to be freed of our flaws. Indeed, we might normally be quite attached to them. This, really, is the problem: how, given the Stoic premise, can we make sense of such perverted attachments?

Perhaps there is not a satisfactory answer to this question within Stoicism. The Stoic frame around Kant's account of radical evil, at any rate, bears traces of intramural struggle with this problem. But the issue closest to the surface in the frame is a shifting attitude to the prospects of moral improvement, given human corruption. As we have just seen, the optimism of the opening frame rests on the canonical idea that we corrupt ourselves; when this optimism goes missing in the closing frame, we might then expect that the Stoic hold on the canonical idea

---

<sup>17</sup> See previous note.

that we corrupt ourselves has proved tenuous in some way as well. In preparation for turning to the closing frame, then, we first need to look into that canonical idea more closely.<sup>18</sup>

## 2.2 The canonical Stoic account of human corruption

What can be said of the origin of human badness, if we accept that nature predisposes us to no fault? As we have noted, the outline of any Stoic answer links our corruption to the acquisition of reason. We need to explain this idea a bit further. To do so, we might begin by considering the reported view of Chrysippus on the matter: “the rational animal is corrupted either by the persuasiveness of external things [διὰ τὰς τῶν ἔξωθεν πραγματειῶν πιθανότητας] or by the influence of associates [διὰ τὴν κατήχησιν τῶν συνόντων]” (DL 7.89). On the face of it, this report seems to say that we are corrupted by external objects and other people. How does it square with the idea that we corrupt ourselves in some manner that is linked to the acquisition of reason? Here the shorthand of Diogenes’ report requires further articulation.

For the Stoics, the link between the human badness and the acquisition of reason also involves a conception of freedom: to say that we bring badness upon ourselves is to say that we do this freely. That is what we need to understand in order to appreciate the canonical Stoic idea that we corrupt ourselves. And to do that, we first need to take a step back and consider some basic points of Stoic psychology.<sup>19</sup> The basic items of Stoic psychology are “impressions” (*phantasia* in Greek, which Cicero renders *impressio*, giving us our word). Impressions that impel an animal, rational or otherwise, to act are called “impulsive” (*phantasiai hormētikai*). Impulse on the Stoic view suffices to motivate action: that is, it will be expressed in action unless something prevents it. A rational impression differs from a non-rational impression, most basically, in its being had by a rational mind — the mind of an adult human being, someone who has acquired reason. Rational impressions are correlated with proposition-like items that Stoics call *axiōmata* or *lekta*. Thus the impulsive impression of an adult human being is the expression of an evaluative point of view, a view about what is worth going for and avoiding, and why. Non-rational impression, by contrast, is just immediate impulse. Finally, the Stoics think that the acquisition of reason completely transforms what can figure as an impulse at all: there can be no

---

<sup>18</sup> As we will see, later Stoics like Posidonius apparently questioned this account of human corruption; hence the qualifier “canonical”, by which I simply mean what is commonly attributed to the “old” Stoics, or otherwise taken to be “orthodox”.

<sup>19</sup> For a particularly helpful overview see Brennan (2003); and for greater detail, Inwood (1985).

impulse now without assent. Its motive force either just is this assent, or else is immediately generated by it.<sup>20</sup>

For our purposes, we need to consider these ideas in relation to the Stoic premise, which is a claim about how we are created. In the positive version, which we just saw invoked by Seneca, it is the claim that we are predisposed, in the way we are created, to the good. As we will see in §3, the Stoics take this claim to hold generally of animals, inasmuch as the good of a creature is the completion or perfection of its nature. An animal is a creature left to its own care, in the sense that it has to *do* certain things — act appropriately — if it is to realise or complete its good. A non-rational animal is created so that some of the impressions that it receives through perception directly impel its appropriate action. As Michael Frede puts it, such a creature “is constructed in such a way that it, by and large, is made to do what it needs to do by the objects which are conducive to its maintenance” (2011, 73). But a human being is an essentially rational animal; the acquisition of reason brings with it resources of language and inferential articulation, as well as some kind of reflective distance. In acquiring reason, a human being is no longer moved immediately by impulses, but acts instead on some conception or understanding of the good. Thus it means something particular to think of an essentially *rational* animal as a creature left to its own care. Nature does not make it do what it needs to do in order to maintain itself and complete its own good: such an animal can only do this on her own initiative.<sup>21</sup>

All of that belongs to the story about how we are created. Strictly speaking, our corruption is not part of that story. Yet by Stoic lights, the fact is that we invariably assent to false evaluative propositions just as soon as we have the resources of reason. Why? When Diogenes says that we are corrupted by “the persuasiveness of external things”, he uses a word (*πιθανότητα*) that is primarily used of persons and arguments, suggesting that, more generally, the origin of corruption is to be traced to the mistaken evaluative commitments of essentially communicative, social creatures. We attach ourselves to things in the world that we take to be of great significance to our lives going well, and we form these attachments in a social context. But if (as for the Stoic) the only true happiness lies in the completion of our rational nature, we must be mistaken about the value of these external things. These ideas are crystallised in the later Stoic Epictetus, who stresses that the only thing that is truly your own is the quality of your own mind, as expressed in the “use” you make of impressions — that is, in acts of assent.<sup>22</sup> Thus when we attach ourselves unreasonably to external things and the values of others, we make

---

<sup>20</sup> For discussion of the debate around this final point see Inwood (1985, 61-63); this debate does not matter for our purposes.

<sup>21</sup> I am thoroughly indebted throughout this paragraph to Frede (2011), especially Chapter 5.

<sup>22</sup> These are persistent themes in Epictetus, but see (e.g.) *Handbook* 1.1 and 1.6.

ourselves dependent on what is not really our own. As Frede puts it: “Then the world begins to have an enormous power over us, and we begin to act under compulsion” (2011, 77). We give over the freedom proper to our rational nature from the first exercise of reason, at least inasmuch as we are not yet prepared to value what is truly our own, which ultimately can only be the perfection of our essentially rational nature in virtue.

### 2.3 The closing frame

The opening-frame Stoic finds grounds for optimism about moral progress in the core Stoic tenet that we corrupt ourselves through chronic, endemic misvaluing. We can undo what we have wrought; we can unlearn our mistakes. The closing-frame Stoic views the prospects of our moral progress in a dimmer light, and we should wonder why. As I will suggest, the closing-frame Stoic has a more tenuous grip on the core Stoic tenet that we corrupt ourselves; and this shift is a reflection of intramural Stoic debates about how to explain the cause of human badness given the premise that nature predisposes us to the good.

In the closing frame (Rel 6:57-9), Kant singles out the Stoics among “ancient moralists” for recognising that moral progress involves something more than letting “the germ of the good which lies in our species develop unhindered”: they recognised that “there is in us an active and opposing cause of evil which is also to be combatted” (6:57). There was no mention of this cause of evil in the opening frame; and as we will see in §3, prominent Stoic discussions of human development (e.g. in *De Finibus* 3) also say nothing about it. So the closing-frame Stoic acknowledges a difficulty that Stoic tradition sometimes fails to acknowledge. This sense of difficulty is further reflected in a conception of moral life as a *battle* against this “cause of evil”. Battles can, of course, be won. But battle is dangerous, and keeping watch for an enemy is exhausting. We are likely to pounce on something harmless, and succumb when the real enemy steals over us unawares. The spirit of the passage is decidedly more pessimistic.

Moreover, after praising the Stoics for recognising that there is a “cause of evil” in us, Kant then charges them with misidentifying it. According to Kant, they looked for it “in the natural inclinations” (*natürliche Neigungen*) when in fact it is “an invisible enemy that hides behind reason and is thereby all the more dangerous” (6:57). It is not clear what Kant could mean by this. Although he does not explain here what he means by “natural inclination”, the contrast he sets up between it and an enemy that “hides behind reason” might appear to suggest that he takes “natural inclination” to be some kind of non-rational impulse. But if he did mean this, he

would be getting the Stoics wrong on a fundamental point.<sup>23</sup> For as we have seen, the Stoics insist that we corrupt ourselves, and that this corruption is expressed in the ordinary emotions or *pathē*, which are a species of rational impulse, since they express assent to evaluative propositions. Curiously, the view that Kant himself endorses here — that the real enemy of moral progress “hides behind reason” and is therefore “invisible” and “all the more dangerous” — looks rather like the Stoic view. What could Kant be saying here?

We might begin by considering the phrase “natural inclinations” — which, again, Kant does not explain. He does consistently (and indeed earlier in the *Religion*) gloss “inclination” as “habitual desire” (*habituelle Begierde*, Rel 6:29).<sup>24</sup> The desire is habitual because it has been repeatedly gratified in a certain manner. But this does not mean that inclinations as such are non-rational impulses. To be sure, non-rational animals and pre-rational humans have inclinations that can presumably be deemed “non-rational”. But by Kant’s lights, to come into the use of one’s reason is to act on maxims, even if one is not deliberately reflective about what those maxims are. Inclinations in such a person will express a commitment to satisfy some sort of need or longing in some sort of way, or with some sort of object. Thus in the *Anthropology*, Kant explains the concept of inclination as “[s]ensible desires that serve as a rule (habit) for the subject” (7:265): it expresses a commitment to a view about how to go on in certain circumstances. Reasoning along similar lines, Allen Wood remarks that for Kant inclination “is already something in which we are rationally complicit and for which we are responsible” (2018, 100). Since Kant does not take “inclination” as such to mean “non-rational impulse”, perhaps we should be wary of supposing that he takes the Stoics to make this identification.

Of course, Kant’s claim concerns *natural* inclinations. What work does this modifier do? If its intended contrast class is *acquired* inclinations,<sup>25</sup> then it might signify that we are predisposed by nature to habitually desire certain things, and when we in fact do habitually desire these things, we have what can be deemed a “natural inclination”. Among the objects of such inclinations would presumably include things the Stoics say are worthy of “selection” for their value in making things go well for one — e.g. health and physical comfort — but not genuinely good, which can only be the completion of our rational nature in virtue. Then Kant would be saying something quite uncontroversial: namely that, by Stoic lights, the enemy of moral progress lies in evaluative confusion, taking what is at best properly indifferent to be genuinely good. But one might not be confused about the value of the objects of “natural inclination” so conceived

---

<sup>23</sup> Baxley (2010, 71-72) fails to recognise this, accepting Kant’s interpretation at face value.

<sup>24</sup> See also Anth (7:251, 265); MS (6:212).

<sup>25</sup> See Ak (20:77-8), from Kant’s notes on moral philosophy, for such usage.

— one might recognise an indifferent as an indifferent — so there should be no identifying the “natural inclinations”, so conceived, as the “cause of evil”.

None of the interpretive options just sketched avoids attributing non-trivial misunderstanding of Stoic ethics to Kant. If one of Kant’s overarching aims in this work is to demonstrate to an audience of eighteenth-century Lutheran theologians the compatibility of certain “fragments” of revealed religion with rational religion (Rel 6:12),<sup>26</sup> then it may well serve his polemical purposes to distance himself from Stoicism, to signal that he is arguing on the side of Christianity. Quite possibly this explains his relatively basic misrepresentation — or misunderstanding — of Stoic ethics in the closing frame.<sup>27</sup> However, if we can avoid charging Kant with basic interpretive confusion, all the better. Nor need we take the shifting portrayal of Stoicism from the opening to the closing frame as a sign of Kant’s woolly-headedness; it might rather be evidence of his appreciation that Stoicism evolved over centuries, developing both in response to its philosophical enemies and to the more earnest pressures of intramural debate. Let us consider this possibility.

Recall that Kant begins by crediting the Stoics with recognising that some “cause of evil” lies in us, and then charges them with misidentifying this cause. This explanatory question was raised within Stoicism in antiquity. The main testimony we have of this intramural debate comes from Galen’s anti-Stoic *On the Doctrines of Hippocrates and Plato*, where Posidonius is said “to censure and refute” Chrysippus for being unable to explain the cause of human badness or vice (PHP 5.5.9/169C).<sup>28</sup> The interpretation of Galen’s testimony is controversial,<sup>29</sup> and we have no evidence that Kant read Galen. However, there are traces of the Posidonian response to the question about the cause of evil in the closing frame’s battle metaphor; and Kant’s use of this battle metaphor is quite plausibly traced to Seneca — indeed, without even needing to stray beyond *De Ira*, the very text Kant quotes in the opening frame. Therefore, even if Kant did not know of the intramural Stoic debate from Galen, he plausibly had some kind of indistinct

---

<sup>26</sup> See note 4.

<sup>27</sup> Regarding the significance of this Lutheran audience for Kant’s handling of his topic, see Wood (2010). Allen Wood suggested to me, in personal communication, that Kant is making a point of distancing himself from Stoicism in this passage (Rel 6:57), “even if he must misunderstand their view in order to do so”. Certainly the closing-frame passage suggests that Kant is not *trying* to signal any allegiance with Stoicism.

<sup>28</sup> Citations of Galen track the passage in the De Lacy edition of PHP, followed by the assigned fragment number in the Edelstein/Kidd edition (Posidonius 1972 and 1999); translations are my own, but I have consulted both De Lacy and Kidd.

<sup>29</sup> Posidonius is sometimes taken, largely on the basis of Galen’s testimony, as an eclectic with strong commitments to Platonic psychology (e.g. Sorabji [2000, 93-120]). The opposing line in Cooper (1999) reconstructs Posidonius’s view in a manner that aims to free it from the polemical ends of Galen’s testimony; I’ve drawn broadly on Cooper in my sketch of Posidonius’s proposal below. See also Kidd (1996 [1971]).

awareness of Stoic responses on the question of “the cause of evil” filtered through Seneca; and this may have provided him with all the fodder he needed to present as “the Stoic” view a position that is not, by any account, entirely orthodox.

In the closing frame, Kant suggests that the Stoic view of moral life as a kind of battle is reflected in “their keyword *virtue* [ihr Losungswort *Tugend*]” — evidently thinking of *virtus*, as it “designates courage and valour (in Greek as well as in Latin) and hence presupposes the presence of an enemy” (6:57).<sup>30</sup> Although by Stoic lights this enemy is manifest in the *pathē* generally, in *De Ira* Seneca deploys the metaphor in regard to anger in particular:

Once it has begun to carry us off course it’s difficult to sail back to safety, since not a jot of reason remains once the passion [adfectus] has been let in and some sovereign right has been granted to it by our own will: it will thereafter do not what you allow but what it wants. The enemy — I stress this point — must be held at bay on the first frontier; when it has entered and made its way through the gates, it accepts no limits from those it has taken captive. (*De Ira* 1.8.1-2)

As we have already noted, Seneca explains the *pathos* of anger as the expression of assent to the complex proposition that one has been done wrong by another, and this ought to be avenged (*De Ira* 2.1.4). Thus we “let” the *pathos* “in” through this act of assent, which evidently means freely ceding “some sovereign right” over our own minds. Once we do this, the *pathos* seems to have a force of its own — that is, something beyond the motive force of the assent itself. (We will return to that point, which is crucial.) Seneca then pauses to say that we should not be misled by the metaphor:<sup>31</sup>

Indeed, the mind is not sequestered, keeping a watch for the passions as things external and apart, so that it can keep them from going farther than they ought. Rather, the mind turns itself into the passion [sed in adfectum ipse mutatur]: that is why it cannot summon back its useful, healthy vigour once it has been betrayed and weakened. (*De Ira* 1.8.2)

On the Stoic view, adult human mindedness is rational through and through — though this rationality is not normally in a state of health or excellence. The *pathos*, then, can only be a corrupted form of rationality, which Seneca expresses by saying that “the mind turns itself into the passion”. The *pathē* are not really “things external and apart”: they themselves are an expression of rational mindedness — corrupted rational mindedness, but rational mindedness all the same. Indeed, when Cicero deploys the metaphor, he speaks more precisely: the *pathē* are

---

<sup>30</sup> Kant cannot be thinking of *aretē*, the general term for virtue or excellence in Greek, but presumably of *andreia* (courage); *virtus* is more plausibly a general term for “virtue” in Latin, though *honestum* sometimes appears to assume this role in Cicero (for discussion, see Graver 2016; cf. Annas 1993, 121n373).

<sup>31</sup> For excellent discussion of Seneca’s use of the military metaphor in this context, see Inwood (2005, 48-51).

manifestations of “a rebellion in the mind as a whole against *right* reason [a *recta ratione defectio*]” (TD 4.22, my emphasis). What is the point of the metaphor? Arguably it expresses something about the experience of ordinary emotion. While any Stoic will say that the motive force of any *pathos* comes from our own act of assent, the metaphor acknowledges that sometimes they at least seem to draw energy from some external source as well.

Posidonius seems to have taken this prospect seriously as a line of response to the question about the “cause of evil”. Posidonius, according to Galen, was looking for a cause of corruption internal to each of us, and proposed that we are led into evaluative confusion “by affective pull [διὰ τῆς παθητικῆς ὀλκῆς]” (PHP5.5.21/169E). Posidonius arrives at this proposal by pressing first on the idea that the evaluative confusion expressed in ordinary emotions is canonically described in Stoic tradition as an agitation of mind, an “excessive impulse”. On Galen’s account, Posidonius challenges Chrysippus to explain “what is the cause of this excessive impulse” (PHP 4.3.4/157F)? Any Stoic account will maintain that the impulse in question must still be a rational one, so that its motive force is the expression of one’s assent to an evaluative proposition. Thus when Posidonius proposes that evaluative confusion arises through “affective pull”, to remain plausibly Stoic he must also maintain that the impulses are rational in the sense that their force is originally the expression of such assent. Yet Posidonius seems to think that assent is not sufficient to explain the excessiveness of the impulse, and suggests that there must be something else involved, something beyond the motive force bound up in this evaluative attitude. As John Cooper puts it, Posidonius suggests there must be “another sort of psychic energy” that can be joined with the rational impulse “so as to increase its force in the ways required to generate emotions” (1999, 466). The result of such a proposal, however, is something that pushes the sort of Stoic optimism that we saw exemplified in Kant’s opening frame to the brink: if we do not *entirely* bring our faults upon ourselves, then we cannot simply undo what we have wrought. Some of the force comes from non-rational resources, on which the Stoic therapy has no traction.<sup>32</sup>

#### 2.4 Significance of the Stoic frame for Kant

Why does Kant insert his account of radical evil within this frame? The opening frame prepares us to appreciate that Kant’s account of radical evil begins with a version of the Stoic premise that nature predisposes us to the good, which we will consider next (§3). Because Kant shares at least some version of this premise, he shares the problem: how do we corrupt ourselves, if we are thus created? But while Kant inherits this problem, he understands it differently. For the Stoics, the

---

<sup>32</sup> Cooper (1999, 481-3) points to this as a consequence of Posidonius’s proposal.

difficulty lies in explaining how reason corrupts itself. That view of the problem leads Seneca to worry, in passing, that it then becomes hard to see what resources a mind thus corrupted could have to restore “its useful, healthy vigour” (*De Ira* 1.8.2): since the mind is just reason, what would remain to restore it to health? However, as we will see, Kant conceives of our corruption differently: it is the “original predisposition to the good” that gets corrupted, not reason itself. This conception presents a somewhat different practical problem: namely, that we will normally be attached to our faults, and thereby unwilling to be freed of them.

Now, the theme closest to the surface in this frame concerns the prospects of moral progress. The optimism of the opening frame is rooted in the Stoic tenet that we corrupt ourselves. This optimism goes missing in the closing frame, and so the closing-frame Stoic appears not to be entirely committed to this tenet. This change in the underlying view about our responsibility for our own corruption is partly expressed in the battle metaphor: the image of an external enemy expresses the worry that there might be forces at work in our corruption beyond our assent. Kant, however, wants to maintain full grip on the idea that we corrupt ourselves: we are entirely responsible for it.<sup>33</sup> So he rejects closing-frame Stoicism. But he also rejects the easy optimism of opening-frame Stoicism. For Kant wants to give the difficulty of moral life its due; and if human badness arises from the corruption of the original predisposition to the good, some genuine difficulty is involved in simply being *willing* to be freed of our faults. Only once we are so willing can we begin to make ourselves good.

### 3. The original predisposition to the good

Kant’s account of radical evil in *Religion* 1 (6:28-44) is preceded by the assertion that there is an “original predisposition to good in human nature” (6:26). With this, I argue, Kant alludes to the Stoic premise that nature gives us uncorrupted starting points. This allusion should not surprise us, for two reasons. First, it is this premise that gives rise to the problem of how — if we are thus created — we corrupt ourselves. This is the problem of the origin of human badness, or

---

<sup>33</sup> “The human being must make or have made *himself* into whatever he is or should become in a moral sense, good or evil” (Rel 6:44). Wolterstorff (1991) and Hare (1996, 60-2) point to this as Kant’s “Stoic maxim”, and see it as raising a problem, from a Christian point of view, for Kant: for it seems that we need help to make ourselves good, and thus Kant wants to appeal to divine assistance or grace, controverting the so-called Stoic maxim. Their “Stoic maxim” is perhaps not entirely misnamed, inasmuch as it is central to the views of both Kant and the Stoics that we make ourselves good or bad through acts of assent. However, if their “Stoic maxim” is meant to express an ideal of moral individualism, then it is miscast either as “Stoic” or as something Kant endorses: both Kant and the Stoics take it that we need help from others to become good, drawing on various conceptions of cosmopolitanism (see Kant on the “ethical community” later in the *Religion* 6:95-101). Kant treats grace as a *parergon* in a border around rational religion: a thinkable possibility, but unfit for any theoretical or practical purpose (6:52-3; see also *Conflict of the Faculties*, Ak 7:55-9).

radical evil. Second, the Stoic premise helps secure the crucial point that we bring this corruption upon ourselves: nature (or Zeus or God) can take no blame for it.

On the Stoic account, nature positively orients us to the good by endowing us with natural affections by which we are attached to our own uncorrupted constitution. This is the Stoic doctrine of *oikeiōsis*.<sup>34</sup> Kant's account of the "original predisposition to the good" accordingly draws in some measure on the Stoic doctrine of *oikeiōsis*. My aim in this section is to establish this link, and identify the important departures.

### 3.1 Stoic *oikeiōsis*

Seneca provides an important account of *oikeiōsis* in Ep 121, which, together with the closely related Ep 124, will help us appreciate the sense in which Stoics conceive of it as, effectively, an "original predisposition to the good". Seneca's account of *oikeiōsis* is part of an attempt to work out what must generally be true of animals, as creatures left to their own care (see Ep 121.18). Taking as given examples of skilful animal action — spiders spinning webs, overturned turtles righting themselves — Seneca argues that any animal must therefore have "a primary attachment to itself" (121.17). It is with reference to this attachment that its actions are *appropriate*, and potentially express an embodied intelligence about what preserves its existence in its own constitution. Seneca thus argues that "all animals have an awareness of their own constitution" (Ep 121.5), and in this a kind of self-awareness. But it is not a self-awareness that, as such, requires the resources of rational thought: it is not propositional or articulate, but "crude, schematic, and vague" (121.12).<sup>35</sup> "If I do everything because I am taking care of myself," Seneca continues, "then care of myself is prior to everything. This care is a feature of all other animals; it is not grafted onto them but born into them" (121.17). With this metaphor, Seneca underscores that this primary self-affection drives the natural growth or development of any animal, as a creature left to its own care.

Seneca scarcely considers the particulars of human *oikeiōsis* in this letter, leaving the protest that he imagines from Lucilius — "What does this have to do with ethics?" (121.1) —

---

<sup>34</sup> The best general gloss of *oikeiōsis* that takes account of the origin of the term may be that offered by A.A. Long: "The Stoics took the word οἰκεῖος, which was the standard word for expressing the relationship between members of households and the ownership of property, and applied it (and related terms) to an animal's affective relationship to itself" (1996, 253). The Latin equivalent is *conciatio*: it figures (e.g.) in Seneca's statement of the Stoic premise at (Ep 94.56) quoted in §2.1 above. *Oikeiōsis* is notoriously difficult to translate; for discussion consider Inwood (1985, 184-6) and Pembroke (1996 [1971]).

<sup>35</sup> As Inwood points out, this can only be "a form of self-perception which does not constitute knowledge in any strict sense" (Seneca 2007, 338). Martin (2015) interprets Ep 121 as a kind of transcendental argument, *from* the given fact of skilful animal action *to* this self-awareness as the condition of its possibility.

dangling. The answer turns on the idea that *oikeiōsis* is conceived as a predisposition to *good*, which Seneca takes up in Ep 124, explaining that “good” in a qualified sense is the completion of a living thing’s nature. In this sense there is even “a kind of good for wheat”, that is manifest only “when the summer rain and the appropriate passage of time have brought the grain to ripeness” (124.11). Something similar is said about the good for non-rational animals — it, too, consists in the completion of their nature — even though they must act appropriately to bring this about. But since the nature of the cosmos is itself rational, by Stoic lights, a rational animal can in principle be complete in an “unrestricted sense” (Ep 124.20) — “complete in accordance with the nature of the cosmos” (Ep 124.14). And as a rational being is “granted the ability to know *why*, to *what* extent, and *how*” (Ep 124.20), Seneca indicates that this completion will require some *understanding* of this unqualified good. The Stoic theory of human *oikeiōsis* is principally concerned with our attachment to this good.

When Cicero presents this theory through the character of Cato in *De Finibus* 3, he accordingly considers the role that acquiring an understanding of the good plays in developing appropriate human attachments. The Stoic theory of human *oikeiōsis* is an account of human development that follows through on the premise that nature gives us uncorrupted starting points. Thus while, on the Stoic view, we inevitably corrupt ourselves as we acquire reason, the Stoic account of human *oikeiōsis* abstracts, in principle, from this corruption. This is why Cato is silent about corruption in his account of human *oikeiōsis*,<sup>36</sup> with a result that is reminiscent of Kant’s opening-frame Stoic: Cato tells the story of the unhindered development of the “seed of goodness” in us.<sup>37</sup>

Cato begins with the idea, which we have already considered in Seneca, that every animal “favours its constitution and whatever preserves its constitution” and likewise “recoils from its destruction and whatever appears to promote its destruction”, and concludes from this that the “primary impulse” of animal movement comes “from self-love” (DF 3.16).<sup>38</sup> As in Seneca, this primary attachment provides the basis of appropriate action (*kathekonta*, *officium*), which Cato explains is a technical Stoic term for action that preserves one in one’s natural constitution (3.20). The human being comes to understand the intelligence of such action, seeing “an order and as it were concordance in the things one ought to do” (3.21). So the human being is not

---

<sup>36</sup> There is only one paragraph on the *pathē*, the evaluative “disturbances” indicative of vice (DF 3.35), at some remove from the first instalment on *oikeiōsis* (DF 3.16-25); I am not here taking up the second instalment on *oikeiōsis* that finds its ultimate expression in cosmopolitanism (DF 3.62-66).

<sup>37</sup> Hence Cato’s announced concern with “natural principles” (DF 3.19, 3.20), to which corruption is external.

<sup>38</sup> The conception of self-love is expressed using the gerundive phrase *se diligendo*, which has the sense of *selecting* oneself above all others.

only capable of appropriate action under the principle of this primary attachment, but also comes, in the normal course of things, to *think of* such actions as appropriate. Moreover, the human being comes to value this order or concordance — the reasonableness of those actions — “much more highly than those first objects of affection” (3.21). Since Cato’s account of human *oikeiōsis* abstracts from the inevitable fact of our corruption, he simply maintains that our appreciation of the appropriateness of these actions leads us to recognise that the “supreme human good, that good which is to be praised and sought on its own account” can only lie in reason, which in turn positions us to realise that this good does not, indeed, consist in merely appropriate actions, but in “moral action and morality itself [*honeste facta ipsumque honestum*]” (3.21; see also 3.58). And so we are led, in some measure by our acquisition of the conception of the good, to a second affinity or orientation: the *acquired* affinity for moral action (*bonesta actio*) or right actions performed from the right motive (*katorthōmata*, 3.24).

It lies outside of our scope to take up the considerable interpretive and reconstructive tasks that Cato’s spare remarks invite. We can restrict our attention to what is most relevant to Kant’s appeal to an “original predisposition to the good”. Cato says that our attachment to right actions performed from the right motive is “not included among our original natural attachments” (3.22). But this acquired attachment is not formed unnaturally, like a graft: for “it too is in accordance with nature [*secundum naturam*]” (3.22). The attachment is acquired because it is only possible with the acquisition of an understanding of the good, which transforms the original attachment to self.<sup>39</sup> In effect, it is *because* one understands that one’s own nature is rational, and the nature of the cosmos is rational, that one comes to see that the original affinity for what preserves one’s constitution is, properly understood, an affinity for rationality in one’s behaviour. In this way, the perspective of right reason reveals the original and acquired natural affinities not to be two fundamentally separate kinds of incentive on action at all. However, as we are about to see, Kant will insist that they must be fundamentally distinct.

### 3.2 Kant’s variation on the theme

Kant analyses “the original predisposition to good in human nature” into three aspects, which he announces as “elements of the determination of the human being [*Elemente der Bestimmung des Menschen*]” (Rel 6:26). With this, he nods to a long-standing 18<sup>th</sup>-century debate about the

---

<sup>39</sup> Frede (1999) points out that DF 3.20-1 is nearly the only source of the Stoic view about how the acquisition of the conception of the good transforms human motivation — i.e. transforms the original affinity for self into the acquired affinity for *bonesta actio*. I am indebted here to his account of this (1999, 90-93). See Klein (2016) for illuminating discussion of some of the interpretive debates regarding the Stoic doctrine of *oikeiōsis*, which I cannot take up here.

*Bestimmung des Menschen*. The focal meaning of the term *Bestimmung* is “determination”, and in that sense the debate concerned *what it is* to be a human being. But in the context of the debate, the term carried the sense of determination to an end, and thus was concerned with the vocation or “destination” of the species. For Kant, these are flip sides of one coin, since the human being is “an animal endowed with rational capacity (*animal rationabile*)” that is called to “make out of himself a rational animal (*animal rationale*)” (Anth 7:321). Our nature is a perfective, something *to be realised* through the cultivation and perfection of our rational capacity.<sup>40</sup> When Kant speaks of a predisposition *to good*, he alludes to this destination, the completion of our nature; and by deeming the predisposition *original*, he indicates that it has to do with how we are created so that we may be so oriented.

The first two elements of the original predisposition to the good are specific predispositions to the “*animality*” and the “*humanity*” of the human being (6:26). The first is a “physical or merely *mechanical* self-love, i.e. a love for which reason is not required” (6:26): it involves the impulse to self-preservation, to sex, to preserve one’s own begotten offspring, and generally to social communion with other human beings (6:26). It is important not to misunderstand Kant here. Any animal (rational or non-rational) will after all have some such “mechanical” affinity for self, and orientation to its own care; thus this element does not, simply as such, require reason. But in a human being this affinity will be taken up into a practical point of view — it will express itself in views about what is worth doing, and in that sense will ultimately be just as “rational” as the rest of the original predisposition to the good. The second element, the predisposition to humanity, does essentially require the resources of reason, since it involves having a conception of how one is faring that is formed by comparing oneself against others. This comparison elicits an impulse to cultivate one’s talents and develop skill, in the interest of receiving regard or esteem from others (6:27). But while these first two elements have to do with our concerned orientation to ourselves as physical beings, the third is a predisposition to “*personality*, as a rational and at the same time *responsible* being” (6:26). This third element is conceived as some kind of concern for ourselves as imputable persons. Specifically, it is “the susceptibility to respect for the moral law *as of itself a sufficient incentive to the power of choice*” (6:27): we are, in other words, constructed by nature so that we can be moved simply by our recognition of what the moral law requires of us.

---

<sup>40</sup> And so Kant takes the *Bestimmung* of our species to consist in “perfection” (7:322) — though these ideas are not unique to Kant. For discussion of the 18<sup>th</sup>-century debate about the *Bestimmung des Menschen*, see Kuehn (2009); for an account of its neo-Stoic context, see Brandt (2003 and 2007). In the background here is Seneca’s point that the human being is not completed by nature (as is the case with non-rational animals and, indeed, gods) but only through self-directed effort or care (*cura*) (Ep 124.14).

Why take this “original predisposition to good” as a variation on Stoic *oikeiōsis*? First, the idea of a “predisposition” just is one of the ways the Stoic doctrine is expressed. Nature does not *predispose* us to any fault, said Seneca: *nature* does not attract us to faults, does not make us regard these faults as properly our own. Nature predisposes us to the good. Unlike Cato, Kant does not abstract from the fact of human corruption when he considers this predisposition. Indeed, he keeps this fact in view by speaking almost as much about the various vices that “can be grafted” onto the first two elements of this predisposition,<sup>41</sup> as he does about the predisposition itself. Of course, the language of “the graft” indicates that this attachment is in some sense unnatural. For a graft is not a natural outgrowth of the rootstock: rather it is fixed on it from outside, so that it can draw the energy of the rootstock — turning it away from its own proper end (*perverting* it) so that it may fix and nourish the scion.<sup>42</sup> Here Kant follows through on Seneca’s conclusion that the primary attachment each animal has to itself is “not grafted onto them, but born into them” (Ep 121.17). Kant accepts this by including self-love as a component of the *original* predisposition to good: it is a necessary endowment in creatures left to their own care; and it is not, simply as such, the source of *our* moral corruption. Kant then says that vice is a graft, to register that it is not natural, even if we inevitably fall into it, and even if it might look and feel natural.

Second, when Kant speaks of three elements of this predisposition, he underscores that *together* they are the predisposition to the good (6:28). His insistence on this point would not make a lot of sense if we assume that “good” means exclusively morally right action and morally good character. But his insistence makes perfect sense if we think of “good” in Seneca’s terms, as the completion of our nature as essentially rational animals.<sup>43</sup> Further, consider how Kant elaborates the point of deeming the predispositions at issue as “*original*”: it is not simply to say that they are born into the species, but indeed to say that “they belong with necessity to the possibility of this being” (6:28). This is exactly the sort of argument that, as I suggested, Seneca

---

<sup>41</sup> These vices include, for the predisposition to animality, “the *bestial vices of gluttony*” and, for the predisposition to humanity, what we could generally think of as anti-cosmopolitan vices, i.e. “vices of secret or open hostility to all whom we consider alien to us” (6:27).

<sup>42</sup> The grafting metaphor is not always used to indicate perversion, either by Kant or Seneca. For if it is supposed that our essentially rational nature is made complete not by nature but by some kind of art and upbringing, then it will be appropriate to speak of the grafting of good dispositions — as Seneca does at Ep 112, and Kant at KpV (5:161). But there are somewhat different senses of “nature” at work in these contexts, allowing for different deployments of the metaphor.

<sup>43</sup> All three aspects count as good not merely on the “negative” grounds that “they do not resist the moral law”, Kant elaborates; rather they are together a positive orientation “*to the good*”, which he explains by saying that “they” — the predispositions — “demand compliance with it [sie befördern die Befolgung desselben]” (6:28). Although the anaphoric anchor for “it” (“*desselben*”) is ambiguous, the most immediate candidate is “the good” (*das Gute*), and not “the moral law” (*das moralischen Gesetz*). Thus the predispositions demand (as it were) that we complete this good, our essentially rational nature.

makes about attachment to self in animals generally: without the “mechanical” sort of self-love at issue in the first predisposition, an animal’s movements would not be a kind of appropriate action expressing embodied intelligence. Such concerned orientation to self is necessary in the very possibility of an animal, a creature left to its own care. We are, of course, considering specifically the rational animal in its “humanity”, and thus as a creature that can freely set ends and act on them.<sup>44</sup> The actions of such a creature are skilful and appropriate in a social context, and require the resources of rational thought. The self-love at issue here is effectively a concern for my own agency — for what I can do, and how I fare as a result — that is judged in comparison to what others can do, and how they fare as a result. There is nothing morally corrupt about self-love of either sort, in and of themselves: they are elements of our inborn affective orientation towards the good, towards our completion as essentially rational animals.

Kant’s absolutely crucial departure from the Stoic doctrine of *oikeiōsis* lies in his handling of the third element of the original predisposition to the good, i.e. the specific predisposition to *personality*. Kant explains this as our readiness to be moved simply by our recognition of what the moral law requires of us: it is, effectively, the endowment of moral feeling. The predisposition to personality is a positive attraction *to* the completion of our nature inasmuch as we are persons, which involves some kind of attraction to morally right action as, in effect, properly one’s own. Now, Cato suggests that human *oikeiōsis* involves two natural attachments: the original attachment to oneself, and the acquired attachment to morally right action, *honestia actio*. On the Stoic view of human development that elaborates this idea of *oikeiōsis*, the latter is a transformation of the first that is possible once one has acquired an adequate conception of the good.

Kant rejects this: our attachment to morally right action cannot be something that arises as a kind of transformation of the attachment to self. They are specifically distinct; one cannot be formed from the other. But both are part of the original predisposition to the good: both are conditions of the possibility of the human being, a creature charged with the completion of its own essentially rational nature. The human being, like any animal, must have an original attachment to self. But the human being can only complete its good by acting freely, on its own initiative. For such actions to promote the completion of its good, they must themselves be manifestations of the autonomy that is proper to a rational being. Thus it must be possible to be moved simply by one’s recognition of what the moral law, the principle of this autonomy, requires. A rational animal not endowed with this possibility would not be capable of choosing to do something simply because his essentially rational nature requires it, and in that sense would

---

<sup>44</sup> This is what “humanity” means, for Kant, as a technical term: see MS (6:392).

not be capable of realising this nature, or completing this good. This, at any rate, is what Kant seems to have in mind when he elaborates, in a somewhat obscure footnote, why the third element of the original predisposition to the good requires its own division (6:26n).

As we are about to see, Kant's account of radical evil explains why the original affinity for self is inevitably perverted, so that it is a corruption of what is *properly* part of the original predisposition to good. He prefigures this account in his remarks here on the general kinds of vice that "can be grafted" onto the predispositions to animality and humanity. But the predisposition to personality, Kant insists, will admit no vicious graft (6:27; see also 6:45). He also takes this predisposition to be just as "original" as the predispositions to animality and personality: hence its existence cannot depend upon the uncorrupted development of the first two. We can see in these claims a reply to Cato, and a reworking of the Stoic doctrine of *oikeiōsis*. If (per Cato) the acquired affinity for morally right action is a transformation of the original affinity for self, and if, on a Stoic view, we inevitably corrupt ourselves with the acquisition of reason, what resources would remain to make ourselves good once we are so corrupted? Yet if the predisposition to the good is genuinely normative, if as Seneca says we are "born to be upright", then it must be possible to restore this predisposition even if we inevitably corrupt it first. In this way, Kant has a clearer view of the original Stoic problem, and what is required for a solution, than the Stoics may have themselves. For Kant, too, is himself trying to think through what must belong to the created nature of a human being, the rational animal left to its own care. This requires, Kant concludes, a specifically distinct predisposition to personality that is incorruptible.

#### 4. Radical evil

As we have just seen, Kant opens his account of radical evil by first endorsing a version of the Stoic principle that nature gives us uncorrupted starting points. This puts his ensuing account of radical evil within the ambit of Stoic debate about the origin of human badness, if we are by nature predisposed to the good. Kant nods to this dialectical framing of the problem at the conclusion of his account of radical evil: since "the original predisposition [...] is a predisposition to the good [...] there is no conceivable ground for us, therefore, from which moral evil could first have come into us" (Rel 6:43). We corrupt the original predisposition to the good, and we evidently do this in an exercise of free choice just as soon as we come into the use of our reason (6:42-3; see also 6:38). But despite claiming that we can make no sense of this corruption, Kant does indeed have things to say about how we do it, and why it is inevitable that we do.

Kant begins by claiming that there is a “propensity to evil in human nature” (6:28), and explains that by “propensity” he means a *liability to be inclined* in certain ways. We have already noted above (§2.3) that Kant takes inclination to be “habitual desire” (6:28). Behind any inclination, then, is a history of gratification: one has gone on in certain ways, in certain circumstances, and will be disposed to continue to do so. In non-rational animals and pre-rational human beings, inclinations obviously involve no commitments about what is worth doing or not doing: there is just an impulse to go on in that way, that is built up over time through repeated gratification. But as we have already noted, for Kant to come into the use of reason is to acquire the resources of a practical point of view: one’s actions are expressions of commitment to practical principles or “maxims”. Behind the inclinations of an adult human being, in other words, is a history of assent to some practical principle, to gratify this sort of desire in this sort of way or with this sort of object. We are rationally complicit in them.

Now, Kant does not think that there is anything “bad” about inclination as such. Whether or not an inclination is “bad” will largely depend on whether it involves assent to a morally bad maxim — or, to put the point in more Stoic terms, whether it expresses evaluative confusion. However, Kant does point to a certain kind of inclination — *Leidenschaft*, commonly translated “passion” — as “without exception *evil* [*böse*]” (Anth 7:267; see also MS 6:408); just as Stoics point to a certain kind of rational impulse — ordinary emotion or *pathos* — as always expressing evaluative confusion. Although there is no scope to consider the connection between Kantian *Leidenschaften* and Stoic *pathē* in any detail, a few points can be briefly noted. Kant conceives of passions as a kind of disease, modelled on cancerous perversion of natural processes from within; or alternatively, he likens them to chains that we cast upon ourselves that then “grow together” with our limbs, impeding natural movement (Anth 7:266-7). Crucially, passions are expressions of rational mindedness: they involve “a maxim on the part of the subject” and are “therefore always connected with his reason” (7:266). They get “themselves rooted” through habitual evaluative thought (7:265): for example, the passion of ambition involves compulsively looking at each situation in light of a commitment to the great importance of being honoured by others. And since Kant takes passions to be “without exception evil”, he surely takes the evaluative commitments driving them to be always mistaken.

So the “propensity to evil in human nature” is a liability to be inclined in ways that express commitment to morally corrupt maxims. To say that such a propensity lies in human nature is to attribute it universally to the species (Rel 6:29); and yet Kant also says, in the very same breath, that this propensity is “contingent for humanity in general” (6:29) — by which he evidently means that we cannot attribute it to the way we are created. For if the propensity is

imputable, and itself evil, then it must be our own deed: it is the original fault, the *peccatum originarium* (6:31). Since Kant takes this deed to be some active corruption of the original predisposition to the good, we should be able to locate it in an account of human development that takes this predisposition as its premise.

The predisposition to animality is a pre-rational orientation to oneself. In a pre-rational state, this orientation does not give us *reasons* for action: for a pre-rational human being acts on inclinations that are not themselves the expression of commitment to maxims. Only when we come into the use of reason does self-love become a *principle* (6:36) — something we *take* as a sufficient reason for action. We are, in this way, poised to act on considerations of self-love just as soon as we are in a position to act for reasons at all. At the same time, Kant reminds us here that the affinity for personality is expressed in moral feeling, the readiness to be moved simply by one's recognition of moral requirement (6:36). So an adult human being can act either on the principle of self-love in the interest of his own happiness, or out of respect for the moral law. There is something in us — something in the way we are created — that inclines or attracts us to both kinds of action. But the stimulus for the predisposition to personality is the exercise of pure practical judgment, which will be unpracticed in someone who is just coming into the use of her reason. Yet anyone at this stage will have had lots of practice with intelligent action guided by the pre-rational affinity for self. Therefore, Kant suggests, the human being's original and default act can only be to take such considerations of self-love to be “*of themselves sufficient* for the determination of one's power of choice, without minding the moral law” (6:36).

Having offered this account, why does Kant insist that “this propensity to evil remains inexplicable to us” (6: 43, see also 6:25, 32, and 41)? What Kant means, in the first place, is that we cannot explain this deed by looking for “its natural cause as an *event* in the world” — in fact, we cannot even coherently think of it as a deed, as something imputable to us, in these terms (6:40). Any kind of causal explanation will entirely miss its target, if this corruption is something we have done, brought upon ourselves. Rather, Kant maintains, we can only coherently think of this deed, our corruption, “as if the human being had fallen into it directly from the state of innocence” (6:41). This sounds something like how Alexander reported the Stoic view: we are “thrown into” badness but “without becoming” so. What Alexander means is that, for the Stoics, the acquisition of badness is not a transition from its opposite through an intermediate state. Rather, the badness is instantaneous. Alexander evidently thinks this is nonsensical, and he perhaps rightly suspects the Stoics of lacking the philosophical resources to elaborate on this idea. But for Kant, this idea is quite naturally expressed: for he distinguishes the causality of freedom from the causality of nature. Our corruption cannot be understood as a transition in

the order of time. For the only thing that it could be a transition of would be a practical point of view; but there was no practical point of view before coming into the use of reason, so it cannot be explained in such terms. Kant then concludes that “we cannot inquire into the origin in time of this deed but must inquire only into its origin in reason, in order thereby to determine and, where possible, to explain the propensity, if there is one, i.e. the subjective universal ground of the adoption of a transgression into our maxim” (6:41). However, while it is coherent to ask after the rational grounds of this deed, such inquiry will not prove enlightening inasmuch as there is no good reason for this choice.<sup>45</sup>

Kant indicates that we are thrown into badness as soon as we come into the use of reason: we are at once entirely bad. Why does he think this?<sup>46</sup> His answer is rooted in his commitments about the metaphysics of the person. If you are to be the imputable source of your actions, there must be some character of your power of choice, something that ultimately moves you. Since we have both the incentives of self-love and the moral law by natural predisposition, this character will be a matter of “*which of the two he makes the condition of the other*” (6:36, Kant’s emphasis). No one is evil because he has uprooted the incentive of the law: Kant thinks this is impossible, incompatible with being human (6:35). The human being is evil only because he “reverses the moral order of his incentives in incorporating them into his maxims” (6:36). This reversal consists in taking oneself to have reason to act from one’s recognition of moral requirement only as long as doing so does not conflict with the demands of self-love. So, since there must be one overarching point of orientation that is the general expression of your causality as an agent, and since this orientation is inevitably at first more fully developed as care for your physical being, self-love will invariably be taken as the condition of compliance with the law. None of this is a causal explanation. Nor is it an explanation from the practical point of view, one that might attempt to explain why this choice makes good sense. It is, rather, a claim about what is inevitable in the development of human psychology, given other more fundamental commitments about the metaphysics of persons.

##### 5. “Restoring” the original predisposition to the good “to its power”

---

<sup>45</sup> This is how Wood (2014, 34 and 2010, 147-9) principally understands Kant’s claims about the incomprehensibility of radical evil: we cannot explain it on rational grounds. However, there *is* good reason for the choice that is required to “restore the original predisposition to the good to its power”, and yet Kant also deems this choice incomprehensible (6:44). He thinks that there is something incomprehensible in any human choice; we will return to this in §5.

<sup>46</sup> The Stoic version of this thought, I take it, is rooted in Socratic intellectualism; but I cannot pursue the issue here.

In the aftermath of this account of radical evil, Kant considers how we make ourselves good, if we are thus corrupted. This discussion comes in a “General Remark” bearing the title: “Concerning the Restoration to Its Power of the Original Predisposition to the Good” (6:44). The restoration of this original predisposition does not *make one good*. Kant rather says that it makes one “by principle and attitude of mind, a subject *receptive* to the good” (6:48). This restoration does not restore the affinity for morally good action full stop (for this affinity can never be dislodged), but rather restores it, as Kant says, *to its power*: so that it might shape or govern an entire practical outlook. (Since corruption is the product of one’s commitment to a certain *relation* between the incentives of self-love and the moral law, in an act of free choice, it is possible for the corruption to be entire, in the sense that it shapes a person’s entire practical outlook, and yet still accommodate an incorruptible element of the original predisposition.) The restoration effectively puts us on the road to becoming good. It puts us in the immediate position to exercise free choice in a way that will enable us to complete our good, perfect our essentially rational nature. But while the corruption and the restoration are each the result of an act of free choice — and thus, Kant thinks, each a complete cast of mind that comes all at once — the actual business of *making oneself good* can only be a matter of “incessant labouring and becoming” (6:48). It can only be done from within an embodied practical point of view, and from that perspective it will involve a gradual alteration in that point of view, through individual acts of choice.

It is widely recognised that Kant in the *Religion* articulates a conception of free choice (*Willkür*) that is arguably only implicit in his earlier works.<sup>47</sup> Attending to Kant’s relation to the Stoics in this context helps us appreciate the significance of this development in a new light. For Kant, in effect, is trying to work out what can be said “within the bounds of reason alone” about divine providence in human creation. Above all, this means that nothing in our created nature can prevent us from making the choices we need to make to complete our essentially rational nature. Obviously this requires a conception of free choice; but the choice must also genuinely lie open to us, to restore this predisposition once it has been corrupted. Therefore, Kant reasons, something in this predisposition must be incorruptible: again, the specific predisposition to personality. Underpinning these claims is a metaphysical commitment that Stoics do not sign onto at all: a dualism between the causality of nature and the causality of freedom. It is this to

---

<sup>47</sup> Kant faced the objection that his conception of the freedom of will (*Wille*) as autonomy might entail that we are not responsible for those actions that are not performed from the motive of duty — on the grounds that only such actions are performed freely, when we are moved directly from our recognition of moral requirement. For otherwise, the objection runs, our actions are the result of the causality of nature, not the causality of freedom; for discussion of the objection, which can be traced to Kant’s contemporary Karl Leonhard Reinhold, see Allison (1990, 133-6).

which Kant chiefly reverts when he suggests that the restoration of the original predisposition to its power is no more comprehensible than its corruption: “How it is possible that a naturally evil human being should make himself into a good human being surpasses every concept of ours” (6:44). By this Kant surely cannot mean that it is incomprehensible on rational grounds, for presumably there is every good reason for the choices involved. Rather, Kant elaborates, there is a change (going from bad to good) that we accordingly can only represent as an event in time and thus as “necessary according to [the causality of] nature”; and yet at the same time must also be “represented under moral laws, as possible through freedom” (6:50).

To understand why Kant deems this incomprehensible, we might briefly revisit the claim about the incomprehensibility of radical evil. Radical evil turns on a choice that is made just as soon as we come into the use of our reason, which is when we acquire the resources of a practical point of view. So there is nothing of the relevant sort to persist through a change, to enable us to think of this alteration as a transition in time. However, the choice that restores the original predisposition to its power is somewhat different, in that this choice is made with the resources of a practical point of view having long been in hand. And yet Kant makes a similar point, claiming that this choice constitutes “a *revolution* in the disposition of the human being” — a “rebirth, as it were a new creation (John, 3:5 [...]) and a change of heart” (6:47). It is a complete change in practical outlook, not some piecemeal alteration of aspects of it. And it is fitting that Kant describes this revolution as bringing on “a change of *heart*”, inasmuch as it restores to its power the original predisposition to the good: with this he returns to the premise of his account of radical evil, that nature predisposes us to no fault. To restore this predisposition from one’s fallen condition involves a complete shift in what one regards as *most dear*, most truly one’s own: morally right action performed for the right reasons, or “merely out of duty” (6:46). The language of the “heart” indicates that Kant’s engagement with Stoic ethics goes well beyond a shared premise: the business of *Religion* 1 is a reworking of the Stoic doctrine of *oikeiōsis* from beginning to end.<sup>48</sup>

Let me connect some dots to elaborate on this last claim. Kant conceives of radical evil as a corruption of the original predisposition to the good, rather than of reason itself. That is a departure from the Stoics. What warrants this departure is, in effect, the worry that crosses Seneca’s mind in *De Ira* 1.8.2 (see §2.3 above): what resources do we have left to restore ourselves to health, if reason itself is corrupted? By Kant’s lights, the corruption of reason itself

---

<sup>48</sup> As an anonymous referee points out, this language registers more familiarly as Christian. My aim is not to contest this, but to show, from another angle, how it relates to the Stoic account of human *oikeiōsis*. There is no reason to suppose, on historical grounds, that these are mutually exclusive observations to make.

would involve the extirpation of its own constitutive principle, the moral law; this he deems “absolutely impossible” (6:35). For the corruption has to come from reason, if we are to be responsible for it; and yet reason cannot annihilate or dislodge its own constitutive principle. Stoic philosophy does not conceive of reason as a faculty with a constitutive principle in quite this way (the prescription to “live in accordance with nature” notwithstanding), and as a result cannot so readily draw a distinction between legislative and executive functions of practical reason, or the will. Kant does draw this distinction, which allows him to suggest that the “original predisposition to the good” gets corrupted rather than reason itself. I am not claiming that this predisposition is itself practical reason, in any form. This predisposition refers to the affective orientation of an essentially rational animal. But this affective orientation is rationally inflected, through and through, once we have come into the use of our reason (see §3.2 above). Kant conceives of this predisposition as a *necessary endowment*. The providential idea here is that the human being has been created in some kind of justice, endowed with everything that it needs to complete its good. When the human being corrupts this predisposition, its actions no longer promote the completion of its nature. The orientation has shifted; the human being acts for some other end. But reason is not corrupted in its constitutive principle, and thus some basic attraction to morally good action — some readiness to regard this as properly one’s own — remains intact. Thus, in effect, Kant responds to Seneca’s worry by reworking the Stoic doctrine of *oikeiōsis* to make it fully consistent with the providential idea that we have been created with everything we need to complete our good.

## 6. Conclusion

In the bit of *De Ira* that Kant quotes in *Religion* 1, Seneca starkly claims that we chiefly need to be “willing” to be freed of our faults, and then nature herself will help restore us. Seneca continues, as we saw, suggesting that it should really be quite easy to find the path to goodness. Kant’s entire account of radical evil, in effect, rebuts this optimism: there is an unacknowledged difficulty in this “being willing”, since our inevitable corruption leaves us quite attached to our faults. When Kant picks up his discussion of Stoic ethics in the closing frame, he elaborates on their error, as he sees it: despite getting so much right — above all, deriving “their moral principle from the dignity of human nature, from its freedom” — they erred in attributing to the human being “an uncorrupted will” (6:58n). Given what I have just argued, by this Kant must mean that they failed to see that we corrupt not reason itself but rather the original predisposition to the good. At any rate, this corruption is the expression of our unwillingness to be freed of our faults.

In the end, Kant maintains that “the Stoic” fails to see that “no matter how far back we direct our attention to our moral state, we find that this state is no longer *res integra*” (6:58n). Maybe this is surprising, since on the face of it Kant is simply saying that we are always already corrupted. However, the Latin phrase alludes to Seneca’s statement of the Stoic premise: “nature does not predispose us to any fault”, but has “begotten us whole [integros] and free” (Ep 94.56; quoted in §2.1 above). I have been arguing at some length that Kant accepts this premise, and accepts the problem of how we corrupt ourselves if we are thus created. Here Kant’s point has to do with the basis on which this premise should be accepted, if we are then to understand the ensuing problem correctly. Our commitment to this premise can itself only be rationally grounded. We accept it because we recognise, as *persons*, that our corruption must be imputable. It’s our deed. But when we think about what *this* means, we should recognise that we can have no epistemic access to our original begotten condition, this *res integra*. There is nothing that such innocence can be like for us, once the resources of a practical point of view — of rationality — are on hand. Since “the Stoic” does not appreciate all of this about the problem, he eventually, in the closing frame, comes to think that our corruption might admit of a causal explanation, as if we had epistemic access to both sides of this altering. Once this move is made, the closing-frame Stoic can no longer maintain a grip on the idea that we corrupt ourselves. But we need that idea, Kant thinks; so we need a conception of freedom that is not available in Stoicism.<sup>49</sup>

---

<sup>49</sup> An earlier version of this paper was presented to audiences at the University of New South Wales and Northwestern University, and I would like to thank the audiences for their questions and comments. I am also grateful to David Bronstein, Markos Valaris, Allen Wood, and Rachel Zuckert for additional discussion and comments. I would also like to thank an anonymous referee and a member of the *EJP* editorial committee for helpful written comments.

## REFERENCES

Abbreviations for frequently cited works are noted in bold below.

### KANT

References to Kant's works follow volume and page of the Academy edition: *Kants Gesammelte Schriften*, edited by the Deutsche Akademie der Wissenschaften zu Berlin (and predecessors), 29 vols (Berlin: Walter de Gruyter, 1900-), abbreviated **Ak**. Quotations generally follow those available in the Cambridge Edition of the Works of Immanuel Kant (abbreviated **CEWIK**), though modified on occasion; the principal volume for present purposes is *Religion and Rational Theology*, translated and edited by Allen W. Wood and George Di Giovanni (Cambridge University Press, 1996). The following abbreviations are also used: **Anth** = *Anthropology from a Pragmatic Point of View*; **MS** = *Metaphysics of Morals*; **Rel** = *Religion within the Boundaries of Reason Alone*.

### OTHER PRIMARY SOURCES

Alexander of Aphrodisias (1991). *Ethical Problems*. (R.W. Sharples, Trans.). Bloomsbury. [**AA**]

von Arnim, H. (2004). *Stoicorum Verterum Fragmenta*. Munich: K.G. Saur Verlag. [**SVF**]

Cicero (1927). *Tusculan Disputations*. (J.E. King, Trans.). Harvard University Press. [**TD**]

Cicero (1998). *De Finibus Bonorum et Malorum*. Oxford: Clarendon Press. [**DF**]

Cicero (2001). *On Moral Ends*. (J. Annas, Ed.; R. Woolf, Trans.). Cambridge University Press. [**DF**]

Cicero (2002). *Cicero on the Emotions: Tusculan Disputations 3 and 4*. (M. Graver, Trans.). University of Chicago Press. [**TD**]

Diogenes Laertius (1925). *Lives of the Eminent Philosophers*. (R.D. Hicks, Trans.). Harvard University Press. [**DL**]

Epictetus (1925). *Discourses*, Books 1-2. (W.A. Oldfather, Trans.). Harvard University Press.

Epictetus (2014). *Discourses, Fragments, Handbook*. (R. Hard, Trans.). Oxford University Press, 2014.

Galen (1981). *De Placitis Hippocratis et Platonis*. First Part: Books I-V. (P. de Lacy, Ed. and Trans.). Berlin: Akademie Verlag. [**PHP**]

Posidonius (1972). *The Fragments*. (L. Edelstein and I.G. Kidd, Eds.). Cambridge University Press.

Posidonius (1999). *Translation of the Fragments*. (I.G. Kidd, Trans.). Cambridge University Press.

Seneca (1762). *Opera Philosophica*. Halle: Impensis Orphanotrophei.

Seneca (1917, 1920, and 1925). *Epistulae Morales*. 3 vols. (R.M. Gummere, Trans.) Harvard University Press. [**Ep**]

- Seneca (1928). *Moral Essays*. Vol. 1. (J.W. Basore, Trans.) Harvard University Press.
- Seneca (2007). *Selected Philosophical Letters*. (B. Inwood, Trans. and Ed.). Oxford: Clarendon Press. [Ep]
- Seneca (2010). *Anger, Mercy, Revenge*. (R. Kaster and M. C. Nussbaum, Trans.). University of Chicago Press.
- Seneca (2015). *Letters on Ethics to Lucilius*. (M. Graver and A.A. Long, Trans.). University of Chicago Press, 2015. [Ep]

## OTHER WORKS CITED

- Allison, H. (2001). *Kant's Theory of Taste*. Cambridge University Press.
- Allison, H. (1993). *Kant's Theory of Freedom*. Cambridge University Press.
- Annas, J. (1993). *The Morality of Happiness*. Oxford University Press.
- Baxley, A. M. (2010). *Kant's Theory of Virtue: The Value of Autocracy*. Cambridge University Press.
- Brandt, R. (2003). The Vocation of the Human Being. In B. Jacobs and P. Kain (Eds.), *Essays on Kant's Anthropology* (pp. 85-104). Cambridge University Press.
- Brandt, R. (2007). *Die Bestimmung des Menschen bei Kant*. Hamburg: Meiner.
- Brennan, T. (2003). Stoic Moral Psychology. In B. Inwood (Ed.), *Cambridge Companion to the Stoics* (pp. 257-94). Cambridge University Press.
- Brooke, C. (2001). Rousseau's Political Philosophy: Stoic and Augustinian Origins. In P. Riley (Ed.), *Cambridge Companion to Rousseau* (pp. 94-123). Cambridge University Press.
- Brooke, C. (2012). *Philosophic Pride: Stoicism and Political Thought from Lipsius to Rousseau*. Princeton University Press.
- Byers, S.C. (2013). *Perception, Sensibility, and Moral Motivation in Augustine: A Stoic-Platonic Synthesis*. Cambridge University Press.
- Cooper, J. (1999). Posidonius on Emotions. In his *Reason and Emotion: Essays on Ancient Moral Psychology and Ethical Theory* (pp. 449-484). Princeton University Press.
- Cooper, J. (2004). Stoic Autonomy. In his *Knowledge, Nature, and the Good* (pp. 204-246). Princeton University Press.
- Frede, M. (1996). Introduction. In M. Frede and G. Striker (Eds.), *Rationality in Greek Thought*. Oxford: Clarendon Press.
- Frede, M. (1999). On the Stoic Conception of the Good. In K. Ierodiakonou (Ed.), *Topics in Stoic Philosophy* (pp. 71-94). Oxford: Clarendon Press.

- Frede, M. (2011). *A Free Will: Origins of the Notion in Ancient Thought*. University of California Press.
- Graver, M. (2016). Honour and the Honourable. In J. Annas and G. Betegh (Eds.), *Cicero's De Finibus: Philosophical Approaches* (pp. 118-146). Cambridge University Press.
- Grenberg, J. (2005). *Kant and the Ethics of Humility*. Cambridge University Press.
- Guyer, P. (2009). The Crooked Timber of Mankind. In A. O. Rorty and J. Schmidt (Eds.), *Kant's Idea for a Universal History with a Cosmopolitan Aim: A Critical Guide* (pp. 129-149). Cambridge University Press.
- Hare, J. (1996). *The Moral Gap: Kantian Ethics, Human Limits, and God's Assistance*. Oxford: Clarendon Press.
- Inwood, B. (1985). *Ethics and Human Action in Early Stoicism*. Oxford: Clarendon Press.
- Inwood, B. (2005). Seneca and Psychological Dualism. In his *Reading Seneca: Stoic Philosophy at Rome* (pp. 23-64). Oxford University Press.
- Kidd, I.G. (1996 [1971]). Posidonius on Emotions. In A.A. Long (Ed.), *Problems in Stoicism* (pp. 200-215). London: Athlone Press.
- Klein, J. (2016). The Stoic Argument from *Oikeiōsis*. *Oxford Studies in Ancient Philosophy*, 50, 143-200.
- Kuehn, M. (2001). *Kant: A Biography*. Cambridge University Press.
- Kuehn, M. (2009). Reason as a Species Characteristic, in A.O. Rorty and J. Schmidt (Eds.), *Kant's Idea for a Universal History with a Cosmopolitan Aim: A Critical Guide* (pp. 68-93). Cambridge University Press.
- Long, A.A. (1968). The Stoic Concept of Evil. *Philosophical Quarterly*, 18 (73), 329-343.
- Long, A.A. (1996). Hierocles on *Oikeiōsis* and Self-Perception." In his *Stoic Studies* (pp. 250-263). University of California Press.
- Louden, R. (2010). Evil Everywhere: On the Ordinariness of Kantian Radical Evil. In S. Anderson-Gold and P. Muchnik (Eds.) *Kant's Anatomy of Evil* (pp. 93-115). Cambridge University Press.
- Mariña, J. (1997). Kant on Grace: A Reply to His Critics. *Religious Studies*, 33 (4), 379-400.
- Martin, W. (2015). Stoic Transcendentalism and the Doctrine of *Oikeiōsis*. In S. Gardner and M. Grist (Eds.), *The Transcendental Turn* (pp. 342-368). Oxford University Press.
- Morgan, S. (2005). The Missing Formal Proof of Humanity's Radical Evil in Kant's 'Religion'. *Philosophical Review*, 114 (1), 63-114.
- Nussbaum, M. (1997). Kant and Stoic Cosmopolitanism. *The Journal of Political Philosophy*, 5 (1), 1-25.

- Pasternack, L. (2017). Restoring Kant's Conception of the Highest Good. *Journal of the History of Philosophy*, 55 (3), 435-468.
- Pembroke, E.G. (1996 [1971]). Oikeiōsis. In A.A. Long (Ed.), *Problems in Stoicism* (pp. 114-149). London: Athlone Press.
- Reich, K. (1939). Kant and Greek Ethics (II). *Mind*, 48 (192), 446-63.
- Schneewind, J. (1996). Kant and Stoic Ethics. In S. Engstrom and J. Whiting (Eds.) *Aristotle, Kant, and the Stoics* (pp. 285-301). Cambridge University Press.
- Sherman, N. (1997). *Making a Necessity of Virtue: Aristotle and Kant on Virtue*. Cambridge University Press.
- Sherman, N. (2013). Moral Psychology and Virtue. In R. Crisp (Ed.), *Oxford Handbook of the History of Ethics* (pp. 744-67). Oxford University Press.
- Sorabji, R. (2000). *Emotion and Peace of Mind: from Stoic Agitation to Christian Temptation*. Oxford University Press.
- Warda, A. (1922). *Immanuel Kants Bücher*. Berlin: Martin Breslau.
- Wolterstorff, N. (1991). Conundrums in Kant's Rational Religion. In P. Rossi and M. Wreen (Eds.), *Kant's Philosophy of Religion Reconsidered* (pp. 40-53). Indiana University Press.
- Wood, A. (2006). The Supreme Principle of Morality. In P. Guyer (Ed.), *Cambridge Companion to Kant and Modern Philosophy* (pp. 342-80). Cambridge University Press.
- Wood, A. (2010). Kant and the Intelligibility of Evil. In S. Anderson-Gold and P. Muchnik (Eds.) *Kant's Anatomy of Evil* (pp. 114-172). Cambridge University Press.
- Wood, A. (2014). The Evil in Human Nature. In G. Michaelson (Ed.), *Kant's Religion within the Boundaries of Mere Reason: A Critical Guide* (pp. 31-57). Cambridge University Press.
- Wood, A. (2018). Feeling and Desire in the Rational Animal. In K. Sorensen and D. Williamson (Eds.), *Kant and the Faculty of Feeling* (pp. 88-106). Cambridge University Press.