

Affects and passions

Patrick R. Frierson

This chapter draws from Kant's *Anthropology from a Pragmatic Point of View* and *Lectures on Anthropology* to develop a Kantian account of the affects and passions in the light of Kant's empirical psychology. In particular, I focus on two key claims about affects and passions from Kant's published writings. First, in his *Metaphysics of Morals*, Kant claims that while affects are merely a "lack of virtue," passions are "properly evil" (MS 6:408, original emphasis). Second, in his *Anthropology from a Pragmatic Point of View*, Kant distinguishes between affects and passions as follows:

Inclination that can be conquered only with difficulty or not at all by the subject's reason is *passion*. On the other hand, the feeling of a pleasure or displeasure in the subject's present state that does not let him to rise to *reflection* . . . is *affect*. (A 7:251, original emphasis)

This passage highlights a couple of important distinctions between affects and passions, most notably that passions are disordered *inclinations* while affects are disordered *feelings*. By providing a psychological account of affects and passions in terms of feeling and inclination, this chapter aims to make sense of Kant's moral assessment of each.

After the first section summarizing changes in Kant's treatments of affects and passions during the twenty years he lectured on the topic, I offer a brief account of Kant's empirical psychology in general. Sections 3 and 4 provide detailed accounts of the psychology of affects and passions (respectively) based on what I take to be his most developed statements about them, and section 5 applies this psychology to the moral assessment of each.

1. Kant's developing views on affects and passions in the anthropology lectures

Two key claims about affects and passions, present in the quotation from the *Anthropology* above, go back all the way to Kant's earliest anthropological

treatments of them: his general definition of affects and passions as states of feeling or desire that preclude reflection and his association of affects and passions respectively with the faculties of feeling and of desire/inclination. Both claims are already present in Kant's earliest lectures on anthropology. The Collins notes from Kant's first course in anthropology (1772–3) lay out his core definition of affects and passions: "A desire that is so big that it makes it impossible to compare the object of our desire with the sum of all inclination, is called affect" (VA-Collins 25:210; see too VA-Parow 25:411, from the same year). And Kant goes on to lay out his key distinction between them, appealing to "an English author," whom he later identifies as Hutcheson (see VA-Friedländer 25:589; VA-Menschenkunde 25:1115), as the source of the distinction:

An English author distinguished, and rightly so, the affects and the passions [*Leidenschaft, oder Passion*]. Passion is a desire, that makes us incapable of seeing the sum of all desires; affect is rather a feeling, which makes us incapable – of consulting the sum of all feelings. (VA-Collins 25:212–13; cf. VA-Parow 25:413)

Both claims persist throughout Kant's lecture courses in anthropology.¹ Despite this apparent uniformity, however, Kant's account of affects and passions changes from his early lectures through his published *Anthropology*.

The first and most striking change is an increasing consistency in distinguishing affects from passions. As the passages cited from Collins make clear, Kant's early lectures, while formally distinguishing affects from passions, fail to remain consistent on this distinction. Thus Kant's definition of affect at VA-Collins 25:210 (also VA-Parow 25:411) identifies *affects* not with *feelings* but with *desires*, precisely the way he later distinguishes *passions* from affects (see A 7:265; VA-Collins 25:212; VA-Mrongovius 25:1339). And this conflation of affect and passion is not a mere accident of these early lectures. In *Parow* (also 1772–3), Kant explicitly says, "In German, one calls affect passion" (VA-Parow 25:412). There, Kant treats *Affekt* as a Latin (*affectus*) or perhaps even English ("affect" or "affection") term, for which *Leidenschaft* (passion) is the appropriate German translation.² Throughout these early lectures, Kant uses "affect" and "passion" as synonyms, and gives examples (such as anger) that he calls both "affect" and "passion."

In these early lectures, then, Kant's introduction of the distinction between affects and passions has something of the importance that a similar

¹ See e.g. VA-Friedländer 25:589; VA-Menschenkunde 25:1115; VA-Mrongovius 25:1339; VA-Busolt 25:1526.

² Brandt (1999), in its note for A 7:251.17–19, points out that the German translation of Baumgarten's *Metaphysica* translates *affectus* in §679 as *Leidenschaften*.

distinction in his initial source – Francis Hutcheson – had. In his *Essay on the Nature and Conduct of the Passions and Affections* (1742, translated into German in 1760), Hutcheson introduces his distinction between affects and passions with the phrase, “When the word *Passion* is imagined to denote anything different from the *Affections*” (Hutcheson (2002 [1742]), 28), a phrase that rightly highlights the casual nature of the distinction in Hutcheson himself. And Hutcheson’s actual distinction between the concepts – that passion

includes, beside the *Desire* or *Aversion*. . . a *confused Sensation* either of Pleasure or Pain, occasioned or attended by some violent bodily Motions, which keeps the Mind much employed upon the present Affair. . . and prolongs or strengthens the Affection sometimes to such a degree, as to prevent all *deliberate Reasoning* about our Conduct (ibid., original emphasis)

– is almost the reverse of Kant’s own. Hutcheson does make an important distinction between desire and mere sensation that is akin to Kant’s distinction between desire and feeling, but Hutcheson’s whole account of affections and passions treats them – as Kant does in these early lectures – as synonymous. And Hutcheson sees neither affections nor passions as precluding reflection in the way that Kant does. Kant seems to have combined his reading of Hutcheson on affects and passions with his own emerging faculty psychology to develop a distinction that he ascribes in these early lectures to Hutcheson, but that is truly his own. In these early lectures, however, Kant follows Hutcheson in being casual about the distinction, making it but then virtually ignoring it throughout his discussion.

Over time, however, the faculty-based distinction between affect and passion becomes more prominent. In the *Friedländer Lectures* (1775–6), Kant continues to conflate affects and passions in certain respects, describing “anger,” for instance, in some places as a passion (VA-*Friedländer* 25:612) and in others as an affect (VA-*Friedländer* 25:599). But Kant develops the distinction in terms of feeling and desire in much greater detail in these lectures. He follows up his introduction of this distinction with an explanation of its implications, noting in particular that passions are oriented towards “what is possible and future” and affects towards “the present,” and Kant uses this distinction to differentiate particular emotions: “Thus fright is a state of feeling. . . therefore it pertains to affect. Longing, however, is a passion. Sadness is an affect. Obsessive ambition is a passion” (VA-*Friedländer* 25:589). And his overall treatment is distinguished into discussions of affects and then of passions, without the general conflation of terms in the previous lectures. In *Pillau* (1777–8), we find very clear statements of the distinct definitions of affect and passion, the former as an incapacity

“to compare a feeling with the sum of all feelings” and the latter as the state “when we lose the capacity to compare an inclination with the sum of all inclinations” (VA-*Pillau* 25:801). Kant experiments with developing a conceptual distinction between “at peace” (*ruhig*) and “content” (*zufrieden*) to distinguish states of being without affect and without passion. And Kant generally distinguishes between examples of each emotional state, though again treats anger as both affect and passion (VA-*Pillau* 25:802). In later lectures, the distinction sharpens, culminating in the clear contrast of *Mrongovius* (25:1339–40), *Busolt* (25:1526) and the published *Anthropology* (A 7:251). The *Busolt* lectures, delivered in 1788–9, go so far as to claim that “where there is much affect, there is little passion, and vice versa” (VA-*Busolt* 25:1526), a far cry from the claim sixteen years earlier that *Leidenschaft* (“passion”) is merely the German term for affect (*Affekt*) (VA-*Parow* 25:412).

Along with the increased emphasis on his faculty-based distinction between affects and passions, Kant also develops further distinctions between the two. Two of the most important developments relate to the different temporality of affects and passions.³ Affects are seen as rooted in the present and of short duration; while passions are oriented towards the future and of long duration. In the earliest lectures, both affects and passions are conceived of as being temporary, even fleeting. Thus the *Friedländer* notes claim, “Both affects [and] passions are an agitation of the mind and not a continual state” (VA-*Friedländer* 25:589). But even within the *Friedländer* notes Kant says, “Some passions are transitory, others persisting,” and then, for examples, mentions that “anger is transitory; hatred, in contrast with it, persists” (VA-*Friedländer* 25:612).⁴ In later lectures (and the published *Anthropology*), the distinction between the transitory and non-transitory will be identified with the distinction between affects and passions.⁵ As the distinction between affects and passions crystalizes, Kant ascribes a different temporality to each: “With desires is not the perception of the actual and present, but rather a presentiment of the future. Feeling relates to the present. True affects belong to feeling, and passions to desire” (VA-*Menschenkunde* 25:1115). But Kant comes to refine this view of

³ Another important difference relates to Kant’s accounts of the natural teleology of affects and passions. In earlier lectures, both passions and affects are seen as provided for by Nature (see e.g. VA-*Friedländer* 25:617), but in later lectures, Kant emphasizes the distinction between affects, which are provided by Nature until reason can take over (see VA-*Menschenkunde* 25:1120, 1123–4; A 7:253) and passions, which are products of social life that are in no cases and in no respect good, but are an unnatural and bad effect of otherwise purposive elements of human nature (our inclinations, our unsocial sociability, and our developing rational capacities).

⁴ Kant even connects this transitoriness of certain emotions with a decreased blameworthiness: “The transitory passions, if they are evil, are sooner pardonable, than the [ones that] persist and have taken root, for these commit bad actions in accordance with rules” (VA-*Friedländer* 25:612).

⁵ See VA-*Menschenkunde* 25:1122; VA-*Mrongovius* 25:1340; VA-*Busolt* 25:1520, 1526; A 7:252–3, 265–6.

each's temporality. For affects in particular, Kant points out that they have an intrinsically future orientation: "Affect can be [rooted in the] present; but its prospect is the future" (VA-*Mrongovius* 25:1343). The difference between affect and passion comes to be tied to the way in which each is oriented towards the future, affect by means of a present sensation that either acts or fades away, passion by means of a fixed interest in future goals.

As these distinctions become sharper, Kant is able to sort different emotional states more clearly into categories. Thus while the earlier lectures see hatred, anger, being in love, avarice, and fear as just several different affects/passions, later lectures come to distinguish sharply between emotions that are properly affects – such as anger, fear, sadness, and pity (see e.g. VA-*Mrongovius* 25:1343–4, 1347) – and those that are properly passions (see VA-*Mrongovius* 25:1356–60). With respect to passions in particular, Kant develops an elaborate taxonomy, within which the passions for vainglory, domination, and greed (VA-*Mrongovius* 25:1356) play particularly prominent roles, along with the sexual/amorous passion (see VA-*Mrongovius* 25:1359).

Kant's lectures on anthropology begin with a general treatment of affects and passions as an undistinguished set of emotions that compromise self-governance by precluding the sort of reflection needed to compare particular feelings/inclination with the sum total of all feelings/inclination. By the time of his published *Anthropology*, Kant maintains this general account but has developed a clear psychological and philosophical distinction between affects – short-term and immediate feelings that overwhelm one – and passions – long-lasting inclinations, consistent with some level of reflection, that dominate one's faculty of desire. In the rest of this chapter, drawing from throughout Kant's lectures where appropriate, I integrate Kant's more developed distinction between affects and passions with his general empirical-psychological account of human action, in order to show how affects and passions work, and why they are ascribed such different moral importance.

2. Kant's empirical psychology in brief

Before turning to the psychology of affects and passions, this section offers some general overview of Kant's empirical psychology.⁶ The central

⁶ Like Kant's particular treatments of affects and passions, his empirical psychology underwent modifications over the course of the time when he was lecturing in anthropology, but this section offers only a brief overview of Kant's eventual empirical psychology.

conceptual framework for Kant's empirical psychology is provided by his distinction between three central human "faculties": cognition, feeling, and desire.⁷ Cognition is a faculty of apprehending objects, whether through the senses, imagination, or reason. Feeling is a subjective faculty whereby one experiences pleasure or pain. And desire is the faculty whereby representations of ends bring about actions directed towards those ends. Given this distinction between kinds of mental state (cognitive, affective, and volitional), Kant explains human actions via interactions between them:

Pleasure precedes the faculty of desire, and the cognitive faculty precedes pleasure . . . [W]e can desire or abhor nothing which is not based on pleasure or displeasure . . . Thus pleasure or displeasure precedes desire or abhorrence. But still I must first cognize what I desire, likewise what gives me pleasure or displeasure; accordingly, both are based on the cognitive faculty. (VM-*Mrongovius* 29:877–8)

Human action is caused by desire, which is caused by pleasure, which is caused by cognition.⁸

Kant further distinguishes higher from lower faculties. Higher faculties of cognition are the rational powers (judgment, the understanding, and reason), while lower faculties are the senses and imagination. Higher feelings are those caused by higher cognitive faculties, such as the feelings of pleasure in maxims to which one is committed. Lower feelings are caused by sensible or imagined awareness, such as the feeling of pleasure in tasting a mango. Desires are higher or lower depending upon the state of the feelings that cause them (i.e. higher feelings cause higher desires). And, for Kant, one explains connections between cognitions and consequent feelings and desires in terms of underlying grounds, such as instincts, inclinations, or – for higher desires – "character."

For the lower faculty of desire, the relevant "cognitions" are sensory, and desires follow from those sensations by instinct or habitual inclination, unmediated by reflection. In contrast, the higher faculty of desire always involves cognition of a practical principle for action and a character that takes up that principle. Even if the cognition of this principle is caused by sensations (direct or imagined), the pleasure and consequent volition are caused by the cognition of the principle rather than directly by those sensations. Thus when one *decides* to "have a smoke," while there may have been an immediate craving that arose from the awareness of certain sensory

⁷ For more detail on Kant's empirical psychology, see Frierson (2005); Frierson (2013); and Frierson (2014).

⁸ See VA-*Friedländer* 25:577; VA-*Busolt* 25:1514; VM-*Vigilantius* 29:1012, 1024.

stimuli combined with an “inclination” to respond to those stimuli with a desire, one’s decision to smoke is based not directly on this craving but upon the taking up of this craving into a practical principle – a maxim – for action: “I’ll have a quick smoke to satisfy my craving.” One who simply finds herself smoking another cigarette without having ever “decided” to do so is motivated by the lower faculty of desire (*mere* inclination).

One important implication of this distinction between the lower and higher faculties of desire relates to the way that feelings prompt each sort of desire. Lower desires are prompted by actually present sensations, so a feeling that prompts direct action-from-inclination is responsive to presently given situations. One takes out a cigarette purely from inclination only in response to a present feeling of pain or discomfort (or a present pleasure at the sight of someone else smoking). But higher desires are responsive to maxims. One who acts on the *maxim* to have a quick smoke can (at least in principle) cognize the principle without the immediate presence of the craving, can plan for future smokes in the light of a principle that covers the future as well as the present. Of course, such a smoker will likely need a present pleasure in the fulfillment of the maxim in order for that maxim to motivate, and, for this particular case, will need to anticipate future pleasures in the satisfaction of future cravings. But the present pleasure is caused by and directed towards a principle that covers more than merely the present. One who smokes *merely* from inclination will, if the present stimulus somehow passes, no longer have any motivation for taking out any cigarettes. A person who smokes from principle can continue to be motivated to act in the light of a principled concern for possible cravings, even while not currently experiencing any cravings.

3. The psychology of affects

Kant describes both affects and passions as “*illness[es] of mind*” (A 7:251) or “emotional agitations” (*Gemüthsbewegungen*) (VA-*Friedländer* 25:589; VA-*Menschenkunde* 25:1115) and classifies them in terms of the faculty of soul that each affects, with affects being disorders of the faculty of feeling while passions are disorders in the faculty of desire/volition (e.g. VA-*Friedländer* 25:589). The disorder common to both is explained by Kant as that through which we “come out of composure”; more specifically, “both affect and passion shut out the sovereignty of reason” (A 7:251). Based on these descriptions, affects and passions would both preclude rational self-governance, and the difference between them would relate to whether

they do this by means of feeling or desire/inclination. In both cases, it is important to distinguish affects and passions from “emotions” and from ordinary feelings, desires, and inclinations. For Kant, even very strong feelings and inclinations need not be affects or passions; they rise to the level of these illnesses of mind only when they preclude reflection or “can be conquered with difficulty or not at all by the subject’s reason” (A 7:251; cf. e.g. *VA-Menschenkunde* 25:1115–18).

But this apparently straightforward account of affects and passions is not sufficient, for two main reasons. First, it leaves unsolved the question why Kant would make such a sharp moral distinction between the two illnesses of mind, calling one “properly evil” and the other a mere “lack of virtue” (MS 6:408). But second, and of more immediate importance, it is not clear precisely *how* affects and passions shut out the sovereignty of reason. And in particular, it is unclear how any illness of mind that is relevant to human actions – as both affects and passions are – could avoid involving *both* feeling *and* desire/volition. Given Kant’s general account of human action, it looks like affects will need to give rise to desires if they are to cause action, and passions will need to involve feelings (and, very likely, disordered ones) if they are to arise at all. But Kant makes clear that while affects and passions “are equally vehement in degree,” “as concerns their quality they are essentially different from each other” (A 7:251; cf. *VA-Menschenkunde* 25:1115). Thus more needs to be said about what precisely is going on in the case of motivation by affects and passions and how this is different from other cases of human motivation. As we will see, getting clearer on how each motivates will also help explain why there is an important moral difference between the two.

We start, in this section, with affects. Kant emphasizes, “it is not the intensity of a certain feeling that constitutes the affected state, but the lack of reflection” (A 7:254). The “reflection” that affects preclude is “the representation by means of reason as to whether he should give himself up to [the feeling] or refuse it” (A 7:251), and in particular a failure to compare “this feeling with the sum of all feelings (of pleasure or displeasure)” (A 7:254; cf. *VA-Menschenkunde* 25:1118; *VA-Mrongovius* 25:1340). Affects are “thoughtless” and involve a sudden “surprise through sensation” that “suspend[s] the mind’s composure,” “mak[ing] reflection impossible” (A 7:252). They arise and dissipate quickly, before one even has time to reflect. Kant compares affects to the “bursting of a dam,” a flash flood (*VA-Menschenkunde* 25:1121–2), a “drunkenness that one sleeps off” (A 7:252), “a fit of madness,” a “strong but temporary whirlwind” (*VA-Menschenkunde*

25:1118) and even a “paroxysm” (A 7:253). Kant also seems to connect affects particularly closely with bodily states, dividing them into those that “*excite* the vital force” and those that “relax the vital force” (A 7:255, original emphasis), describing fright as “dependent for the most part merely on bodily causes” (A 7:256), emphasizing physiological features of affects like anger and shame (e.g. at A 7:260), and even devoting a section to “Affects by which Nature Promotes Health Mechanically” (A 7:261).⁹

Given their particular connection to feeling, affects might seem to have no particular *motivational* import at all. And sometimes when Kant discusses affects, he focuses on them merely as feelings, without regard for motivational efficacy. Thus Kant illustrates the absence of reflection with a rich man who sees a goblet broken and “gives himself over completely to this one feeling of pain (without quickly making that calculation [of the importance of the goblet relative to other goods] in thought)” and thereby “feels as if his entire happiness were lost” (A 7:254; cf. VA-*Pillau* 25:801; VA-*Busolt* 25:1116). Similarly, “fright” seems to be an affect that merely “disconcerts the mind” without implications for desire or action (A 7:255; cf. VA-*Friedländer* 25:589). These problems are described purely at the level of irrationally ordered feelings, without necessary reference to motivational effects.

But, for Kant, affects are not generally motivationally vacuous. Within Kant’s empirical psychology, feelings generally cause desires. And Kant emphasizes that while “there are affects, that directly hit only at sensibility, [there are] others that, besides the senses, also penetrate the soul [that is, the faculty of desire]” (VA-*Menschenkunde* 25:1125). Typically, those with affect “act irrationally” (VA-*Mrongovius* 25:1340, emphasis added) and affect can “double all [one’s] powers” and serve “as a spur to activity” (VA-*Mrongovius* 25:1343). Thus “*Anger* . . . quickly stirs up powers to resist ill” (A 7:255, original emphasis), and “affects” can be “violent,” as when someone “is so angry that she has turned the whole house upside down” (VA-*Friedländer* 25:620–1). The effectiveness of affects at generating activity even leads Kant to say that “the human being manifests greater strength in [a state of] affect . . . than if he is cool-headed” (VA-*Friedländer* 25:615) so that “some people even wish that they could get angry, and Socrates was doubtful as to whether it would not be good to get angry at times” (A 7:253). More crudely, “the affect of fright [can] produce a scream” (VA-*Friedländer* 25:600), and even affects that primarily paralyze rather than stimulate (such as shock and certain cases of fear or anger; cf. VA-*Friedländer* 25:591–2) have direct

⁹ Borges (2008) emphasizes this physical–biological component of affects.

effects – even if only negative – on action. Given that Kant sees most feelings as practical, he quite reasonably moves from discussions of the disorder of *feeling* involved in affect to the ways in which this disorder affects desires and thereby action.

The way affects can be disorders of feeling and relevant to action can be understood by analogy with “temperaments of feeling,” which directly influence feeling but give rise to characteristic actions, as when the sanguine person “attributes a great importance to each thing for the moment, and the next moment may not give it another thought” (A 7:287–8) and as a result “makes promises in all honesty, but does not keep his word” (A 7:288). Because feelings are typically motivational, the character of one’s faculty of feeling affects one’s actions. In the case of affects, sufficiently strong feelings involve a “suspension of composure” (A 7:252; cf. VA-Friedländer 25:589–90, 611), where to “keep one’s composure means . . . the mind is subject to our power of choice” (VA-Friedländer 25:589). The way affects preclude reflection is to compromise the influence of the power of choice; that is, the higher faculty of desire. For affects with volitional importance (whether through provoking actions or paralyzing one’s capacity for action), affects prompt “actions” through bypassing choice: “in affect, the person cannot carry out a rational choice” (VA-Collins 25:212). In terms of Kant’s psychology, one’s actions are motivated solely according to the lower faculties of sensation, feeling, and desire.

As feelings so overwhelming that one that one cannot properly assess their place in one’s overall happiness, affects become *immediate* causes of action, bypassing consultation with higher faculties of cognition and desire. Thus in *Anthropology* (and throughout his lectures), Kant emphasizes that affect relates to “the subject’s present state” (A 7:251; but cf. VA-Menschenkunde 25:1125; VA-Mrongovius 25:1343). Consistent with this emphasis on the lower faculties, Kant points out that affects move through merely animal rather than distinctively human forms of volition.¹⁰ Affects are tied to a “propensity to sink back into animality” (VA-Menschenkunde 25:1125):

Instinct drives only savage people, as long as they are still half animal . . . This affect springs out of a natural instinct and rules us for a while until we are ruled by reason . . . Then the instinct must cease. (VA-Menschenkunde 25:1114)

¹⁰ Accordingly, in his *Friedländer* lectures, where the distinction was not yet as sharp as in later lectures, Kant associates both affects and passions with animality (see VA-Friedländer 25:616–17). Later (e.g. VA-Mrongovius 25:1360–1) Kant emphasizes that passions, unlike animal instincts, depend upon socialization and culture and hence are distinctively human.

Nature [gave] the predisposition to [affects] to our animality... If the human being has emerged from animality, he does not need the affects anymore and must suppress them. (VA-*Mrongovius* 25:1342–3)

Affects are disorders of feeling that so displace humans' abilities to reflect and reason that we sink to the level of animals, either paralyzed with strong feelings or motivated in merely animal ways, without choice or conscious deliberation.

One important challenge to this account is what we might call the “challenge from Kantian freedom.”¹¹ Kant is widely taken to hold a view of human agency according to which human beings are incapable of being compelled by sensuous incentives unless those are freely endorsed. The most common version of this point is framed in terms of the “incorporation thesis,” which takes Kant’s claim that “freedom of the power of choice has this characteristic, entirely peculiar to it, that it cannot be determined to action through any incentive *except so far as the human being has incorporated it into his maxim*” (R 6:23–4, original emphasis) as a general principle of human action. Maria Borges,¹² for example, claims, “As strong as emotions [which in the context particularly includes affects] can be, and as much of a problem for morality as they can portray, the very idea of practical reason presupposes that agents can decide how to act” (Borges (2004), 157). Such a conception of human agency would preclude an account of affects as altogether *bypassing* choice based on maxims (the motives of the higher faculty of desire).

In fact, however, this challenge from Kantian freedom does not pose real problems for Kant’s account of affects. Most of the passages in which Kant seems to preclude actions caused independently of maxims or choice in fact make a narrower point. The classic formulation of the incorporation thesis, for example, is specifically described as an account of how *the power of choice* is determined to action.¹³ But the power of choice is precisely a

¹¹ There are other challenges as well. One, which I call the “challenge of rational affects,” arises from the fact that some affects (notably but not exclusively enthusiasm) seem caused by ideas of reason and hence grounded in higher faculties (cf. Sorenson (2002); Clewis (2009)). Another problem, dealt with in part in the next section, is that my account here implies that affects have their ground in inclinations, in apparent contrast to Kant’s explicit identification of inclination with passion. I discuss both in more detail in Frierson (2014).

¹² Borges is a particularly relevant example here because she is generally very resistant to overemphasizing the role of freedom in governing emotions, specifically targeting the view of emotions laid out by Marcia Baron in Baron (1995).

¹³ Allison, consistent with the claims I make here and throughout this section, is careful to describe the incorporation thesis as a thesis about “rational agency” (e.g. Allison (1990), 5, 40), not about human actions altogether.

power of the *higher* faculty of volition. Insofar as affects bypass this faculty, they also bypass any need to be incorporated into maxims. One might read this passage as a general claim about all action, but the passage itself is narrower in scope. Similarly, Borges's key text in defense of her application of the incorporation thesis to human action in general is taken from Kant's lectures on ethics:

Can I really conceive of a pathological compulsion in man as well? Truly I cannot, for freedom consists in this, that he can be without compulsion in the pathological sense; nor should he be compelled in that way. Even if a man is so constrained, he can nevertheless act otherwise. (VMO-*Mrongovius* 29:617)

But this passage, too, has a narrower application. Kant makes this claim in the context of an example of "fear of punishment" as a "compulsion [that] is pathological" (VMO-*Mrongovius* 29:617), and Kant's point is that in ordinary cases, impulses do not literally "compel" one to act but merely provide a strong but resistible incentive. That does not imply that there cannot be cases in which human beings act directly on feelings of certain sorts – affects – without the reflection required to be "capable of doing otherwise" in any meaningful sense.

In this respect, Kant's classic example of affect – anger – can be misleading. Sometimes feelings of anger literally overpower one, bypassing reflection, but at other times feelings of anger merely exert a particularly strong influence on deliberation. Flinging something at someone in a rage is quite unlike deciding to hurt someone because one is angry at them. That affects apply to the former example rather than to the latter is particularly evident in the context of what Kant calls "wild affects," where the "affect . . . negates its own natural effect" (VA-*Friedländer* 25:591):

For example, one sees a child fall into the water, who one could save, however, through a small aid, but one is so shocked that one thereby cannot do anything. Shock anaesthetizes someone such that one is thereby unable to do anything at all. Similarly, one can be completely shocked by joy over an unexpected good fortune, and indeed also in this way, that one is completely limp; whereas joy should surely, on the contrary, have good consequences, but since the affect is wild, it itself negates its effect. It is just the same with the affect of anger. Anger should, after all, have the effect of taking someone to task and reproaching him, yet often the angry person is . . . is irritated, quivers and trembles, and cannot say a word; that is an unrestrained affect. (VA-*Friedländer* 25:591–2)¹⁴

¹⁴ See too A 7:260 f.; VA-*Collins* 25:211; VA-*Parow* 25:412; VA-*Menschenkunde* 25:1116.

In all three cases, the relevant affect is a strong feeling that prevents reflection, and in each case it is clear that the way in which it prevents reflection is not by misleading deliberation but by forestalling it altogether. These are, as Kant indicates, special cases. But they are special not in the way that they forestall reflection but in the effects of that forestalling. The affect of anger that provokes one to do immediate harm to its object precludes reflection just as much as the paralyzing “unrestrained” or “wild” affect. The difference is that the flinging anger accomplishes its natural effect, while the quivering anger works against that effect.

Given this account of affect, it should be clear that Borges and other interpreters are wrong to think that, for Kant, emotions are always capable of being overridden by practical reason. But it should also be clear that *this* sort of emotional lack of control is not a fundamental problem for Kant’s overall account of human agency. We might say of affects what Edward Hinchman has said of cases where one “is gripped by an arational force”: “There is nothing philosophically perplexing . . . about compulsive action . . . The mental activity or behavior in question simply does not qualify as choice, intention, or action” (Hinchman (2009), 407–8). More precisely, human “actions” motivated by affects are not the actions of humans qua rational agents. While they may still be “intentional” in the sense that there can be a representation of an end that brings about movement towards that end – as in the case of furious rage – they are not “intentional” in the *rational* sense; that is, no end has been incorporated into a maxim that provides a motive for the higher faculty of desire. Thus there is no “choice” in these cases, in either the contemporary or the Kantian sense. This solution, of course, leaves unsettled important issues about the extent to which human beings can be held responsible for affect-driven actions, but I reserve discussion of those issues for section 5.

4. The psychology of passions

Turning from affects to passions, Kant’s account might look superficially similar. Like affects, passions are “illnesses of mind” that “shut out the sovereignty of reason” (A 7:251), and just as affects prevent the comparison of one feeling with others, a passion is an “[i]nclination that prevents reason from comparing it with the sum of all inclinations in respect to a certain choice” (A 7:265).¹⁵ But unlike affects, “the calm with which one gives oneself up to [a passion] permits reflection and allows the mind to form

¹⁵ Cf. VA-*Menschenkunde* 25:1141; VA-*Mrongovius* 25:1339–40, 1353; VA-*Busolt* 25:1519.

principles” (MS 6:408, cf. A 7:266). Thus while passions “can be conquered only with difficulty or not at all by the subject’s reason” (A 7:251), they nonetheless seem to involve reflection to a considerably greater degree than do affects.

To figure out to what extent passions can involve reflection, it is important to clarify in what sense passions are “inclinations.” Kant uses the term “inclination” in two crucially different senses throughout his empirical psychology (and moral philosophy). On the one hand, an inclination in the strict sense is a ground of the lower faculty of desire, whereby certain sensations are connected with volition. In this context, “inclinations” are distinct from instincts in being acquired, and distinct from character in that they relate to the lower rather than higher faculties of feeling and desire. But on the other hand, Kant often uses the term “inclination” to refer indirectly to practical principles that determine one’s higher faculty of desire to pursue ends set by inclinations in the strict sense. Here an “inclination” can refer to any particular practical principle that has sensuously given ends, whether these are given by instinct or by inclination, and it contrasts with *pure* practical principles (the moral law). One who incorporates the end of an inclination in the strict sense into a maxim for action has an “inclination” in this second, derived sense. In the context of passions, one must discern which sort of inclination a passion is.

When Kant identifies passion as an inclination, he primarily has in mind the second sense, which allows passions to involve commitments to principles that make objects of inclinations their ends: “the calm with which one gives oneself up to [a passion] permits reflection and allows the mind to form principles” (MS 6:408). More specifically, “Passion always presupposes a maxim on the part of the subject [and] is therefore always connected with his reason” (A 7:266, see too A 7:410).

Kant’s descriptions of the passionate man even make him sound like a paradigmatic case of true character, since he is consistent and even principled in pursuing his passion. The “cold passions,” which Kant identifies with “*manias for honor, dominance, and possession*” are not only “not connected with the impetuosity of an affect” but *are* connected with “the persistence of a maxim established for certain ends” (A 7:268, original emphasis).

Kant must, then, be distinguishing between two different senses of “reflection” and “principles” here, where passions preclude one sort but allow another. Elsewhere, Kant makes clearer what he has in mind. As in the case of affects, passions involve a lack of *comparison*: “Inclination that prevents reason from comparing it with the sum of all inclinations

in respect to a certain choice is *passion*” (A 7:265). In particular, passion makes a person “blind to . . . purposes which his [other] inclinations also offer him[, which] he ignores completely” (A 7:266). But unlike affects, “passions can be paired with the calmest reflection” and thus “are not thoughtless; rather, they take root and can even coexist with rationalizing” (A 7:265).

Iain Morriison has helpfully characterized a distinction between maxims of different sorts that can helpfully be applied to the case of passions:¹⁶ Some “maxims actually justify actions twice over. They justify actions both in terms of the immediate end contained in the maxim and in terms of the end of happiness” (Morriison (2005), 82). Other maxims, however, justify actions only in one sense; that is, merely in terms of proposing good means to achieve the end contained in the maxim (an end for which one has an inclination). Passions would allow the second sort of maxim – one justified in terms of the immediate end of inclination – but preclude the first – one also justified in terms of overall happiness, or, more generally, a consideration of all of one’s ends (pragmatic and moral).

Thus, to take one of Kant’s favorite examples of a passion – passionate vengeance (see A 7:270) – the maxim “revenge is a dish best served cold” (i.e. “retaliate for wrongdoing only after waiting”) might be well justified in terms of the inclination (passion) for revenge, but might *not* be justified in terms of one’s overall long-term happiness. One with a passion for vengeance would be motivated by this principle, structuring decisions, formulating subordinating maxims, and so on, all in accordance with the desire for revenge. By contrast, one with a pure inclination, *in the strict sense*, for revenge would not even formulate maxims but would simply strike out in retaliation. While this might be possible, Kant would classify it under the affect of anger rather than the passion of vengeance with its lasting maxims. And for yet another contrast, for one with an *inclination* to revenge incorporated through reflection into an overall principle of self-love in the pursuit of happiness (with or without a moral proviso), maxims of revenge would be considered not only in terms of the benefits of satisfying the inclination for vengeance, but also in terms of its effects on one’s long-term well-being. One *might*, in that context, end up endorsing the maxim of patient and fierce retaliation, but one might just as easily – or even more easily – endorse something like the Hobbesian maxim to “in revenges . . . look not at the greatness of the evil past, but the greatness of

¹⁶ Morriison does in fact apply this to the case of passions at the end of his article (see Morriison (2005), 85–7). I largely agree with Morriison’s account, but I think that it conflates passions and weakness of will with insufficient attention to the variety of ways human willing can go astray.

the good to follow. Whereby we are forbidden to inflict punishment with any other design than for correction of the offender, or direction of others” (Hobbes (1660), Chapter 15).

We can make two further refinements to this account of passions. First, while Kant often focuses on the way in which passions preclude reflection *on other inclinations*, passions that prevent reflecting on other inclinations or on happiness as a whole also prevent reflection on requirements of duty (cf. VMo-*Collins* 27:368; MS 6:408–9). Second, passions preclude reflection not by *bypassing* higher faculties, nor merely by outweighing other relevant concerns. The passionate person rationally deliberates, but only in terms of his guiding passion, so nothing unconnected with that passion gets a hearing. But this passion is one upon which the passionate person has settled *intentionally*, and the maxims for the satisfaction of this passion have become abiding principles of the person’s character. One with passion has a principled commitment to pursue the ends of passion, without regard to any other ends.

5. The moral status of affects and passions

The model developed in the previous sections lays out two ways human beings act against their own best interests (whether those be moral or prudential). Affects are disordered feelings that bypass the higher faculties altogether, while passions are disorders *of* the higher faculty whereby it focuses its practical principles around a particular end but does not reflect on the value of that end relative to others. Thus an angry rage wherein one lashes out would be an affect. A hateful vengeance whereby one organizes life-principles around the desire to do harm to another, and does so without considering the moral or prudential cost of these principles, would be a passion. This model makes sense of many of the characteristics that distinguish affects from passions, such as the tendency of affects to arise and pass away quickly and of passions to persist and fester (e.g. A 7:252), but among the most important advantages of this model is that it provides a psychological background from which to understand Kant’s assessment of the moral status of affects and passions. In particular, this model helps make clear why Kant would insist, as he does in the *Metaphysics of Morals*, that affects are merely a “lack of virtue” while passions are “properly evil” (MS 6:408).

For explaining this moral assessment, it is important to clarify precisely in what Kant considers moral evil to consist. When Kant turns to evil in *Religion within the Boundaries of Mere Reason*, he explains,

The difference, whether a human being is good or evil, must not lie in the difference between the incentives that he incorporates into his maxims (not in the material of the maxims) but in their *subordination* (in the form of the maxim): *which of the two he makes the condition of the other*. (R 6:36, original emphasis)

The “incentives” of which Kant speaks here include the moral law and various sensuous incentives, which latter are incorporated into particular maxims by means of a general principle of self-love. Without going into all of the details of his account,¹⁷ Kant’s general point is that to be morally good is to be such that one’s maxims of choice are structured in such a way that the moral law is prioritized over all other practical ends. Moreover, the relevant subordination here must involve *consistent* prioritization of the moral law: “The statement, ‘The human being is *evil*,’ cannot mean anything else than that he is conscious of the moral law and yet has incorporated into his maxim the (occasional) deviation from it” (R 6:32, original emphasis). One can be evil while still subordinating *some* non-moral ends to the moral law, unless one’s character *always* prioritizes the moral law.

In this context, the evil of passions should be clear. A passion is a deliberate orientation of the higher faculty of desire towards promoting the end of a particular inclination. An agent influenced by a passion has a determinate character constituted by consistent maxims in pursuit of a particular good. Because these are maxims of the higher faculty, this constitution of character is ascribable to a free intelligible character.¹⁸ And these maxims are endorsed independently of any consideration of their moral or prudential costs. But human evil is identical to the subordination of the moral law to non-moral incentives, so one with a passion is evil. Nonetheless, passions are a special case of human evil. Generally, evil is understood in terms of the subordination of the moral law to principles of self-love or one’s own happiness. In the present case, even self-love is subordinated to a particular inclination. But the general characteristic of human evil – the prioritization of non-moral to moral incentives – is wholly operative. And since passions work *through* the higher faculty rather than *around* it, one can be held fully responsible for one’s passions. One characterized by maxims that prioritize the ends of a particular inclination to all others – including moral ones – is properly evil.

¹⁷ For more, see Frierson (2003), 108–14; and Frierson (2013), 72–81.

¹⁸ For a defense of the connection between the empirically evident higher faculties and the ascription of deeds to a free intelligible character, see Frierson (2008).

Prima facie, the moral status of affects is equally clear. As operations of the lower faculty of desire, affects bypass humans' power of choice. Thus what one does under the influence of affect is not "action" in the strict sense, and one cannot be held directly accountable for it. Even before Kant had clearly distinguished affects from passions, when he still allowed for both passions and affects to be "[s]ome . . . transitory, others persisting," he has already foreshadowed his insight that the "transitory passions, if they are evil, are sooner pardonable, than the [ones that] persist and have taken root, for these commit bad actions in accordance with rules" (VA-*Friedländer* 25:612). Over time, persistence would be more strongly linked with the rule-governedness of the higher faculty of desire and made a central feature of *passions*, and the transitory and merely animal-like motivational force of the affects would become not only "less blameworthy" but wholly without evil in the true sense. For that reason, affects are a mere "lack of virtue" and, as it were, something childish and weak, which can indeed coexist with the best will" (MS 6:408, original emphasis). Because they bypass the higher faculties altogether, we might say that affects are an *absence* of free agency rather than a misuse of it. What one does under the influence of an affect cannot be ascribed to one's intelligible character and thus cannot strictly be imputed. Put another way, there is no "practical perspective" on actions from affect. One finds oneself to have done (or to be doing) something, which one promptly regrets and rejects, but under the influence of affect, one does not "act" in any sense that implicates a posture of freedom.

However, Kant's account of the moral status of affects is more complex. First, Kant insists that affects, like passions, must be resisted by a moral apathy. In his *Metaphysics of Morals*, affects are introduced along with passions as something that needs to be "subdue[ed]" in order to "be . . . one's own master in a given case" and thereby have the "inner freedom" required for virtue (MS 6:407). And in corresponding lectures, Kant emphasizes, "Anger, to be sure, is also contrary to the duty of apathy, whereby we must not abandon ourselves to any affect" (VMo-*Vigilantius* 27:687). Kant even says, "we blame ourselves, when we let ourselves come into affect" (VA-*Menschenkunde* 25:1118; cf. VA-*Busolt* 25:1527). Relatedly, Kant often treats affects as feelings over the origin of which one has a certain degree of control. He describes one who "lets . . . lively sympathy . . . rise into an affect" (MS 6:408–9, emphasis added), and his account of the man who obsesses over the shattering of a "beautiful and rare crystal goblet" describes him as one who "gives himself over completely to this one feeling of pain" (A 7:254, emphasis added). Both cases involve a reference to something like a free act, and thus some sort of responsibility for the emergence of

the affect. Combined with the obligation to develop an apathy that could prevent and subdue affects, this seems to open room for considering affects not merely a “lack of virtue” but a morally culpable lack of virtue, thus something blameworthy.¹⁹

With respect to Kant’s seeming affirmation that one can be, to some degree, morally responsible for affects, we need to distinguish between moral responsibility for actions motivated by affects, and responsibility for the affects themselves. Kant’s imagery of affects as like the “bursting of a dam” (A 7:252) is apt. Occasionally, as when one’s child falls into the water or one experiences a sudden and shocking harm or insult, affect comes over one like a flood that overtops the dam in an instant: “All affects surprise us, but some surprise us so suddenly that we cannot prepare ourselves for it in the slightest” (VA-*Mrongovius* 25:1342). But often affects rise in a swell, and it is only by “abandoning oneself” or “giving oneself over” that these rising feelings become uncontrollable. Even when this abandonment is largely passive, a matter of simply letting feelings slip out of control, it is still something that we could have done something about: “We are blameworthy when we let ourselves come into the throes of affect; but when we are already in it, we are not capable of pulling ourselves out of it and then are not blameworthy” (VA-*Mrongovius* 25:1342). Moreover, there are longer-term strategies that one can employ to make oneself less susceptible to affects, practicing strategies of calming and refusing to cultivate a heightening of those emotions that are already most likely to lead to affect. Even if there is no responsibility for what one does when one is overpowered by affect, there are many ways that one can maintain and promote self-control before that point. And since all affects undermine the capacity for self-governance, one should constantly strive to eradicate them in one’s life. Insofar as one

¹⁹ A further complicating aspect of Kant’s account is even more troubling, since Kant sometimes seems to suggest that affects, at least those considered “rational affects,” can be morally *praiseworthy* (cf. Kant, ‘Essay on the Maladies of the Head’ 2:267, KU 5:272, SF 7:86; Sorenson (2002), 121; Clewis (2009), 170). However (contra Sorenson), Kant does not endorse any rational affect as morally required or even morally recommended. His apparent ambiguity is due to the fact that this particular affect is a sign of a moral predisposition in human beings and can thus provide an antidote to extreme pessimism about the possibility of morally good action. The presence of rational affects (particularly enthusiasm) indicates humans’ moral predispositions and thereby enduring capacities for virtue, but enthusiasm does not represent a genuine choice to act in accordance with moral ideas and thus cannot be considered morally praiseworthy. While “enthusiasm . . . *seems* to be sublime,” in fact “it cannot *in any way* merit a satisfaction of reason” (KU 5:272, emphases added; cf. MS 6:409); it is “fraught with danger” (SF 7:85) and “deserves censure” (SF 7:86). In fact, “to the extent they turn into affect . . . the noblest agitations of the mind are the most harmful” (VA-*Friedländer* 25:591) since even if “an affect . . . is directed to something good, then [it is] not yet thereby excused, for [it] then must also be constituted this way according to the form” (VA-*Friedländer* 25:591; VA-*Menschenkunde* 25:1119).

fails in that striving towards virtue, one deserves censure and can be held responsible for moral failing. When they become a “lack of virtue” with which one is complacent, affects cross the line into moral evil.

6. Conclusion

Affects and passions inhibit moral action and even prudence. Affects interfere with proper willing by bypassing the higher faculty of desire altogether, as in cases such as shock and outbursts of rage. In that sense, they are intense versions of relatively ordinary non-moral motivation, as when we do something by habit. They are “merely” a lack of virtue, but this is no small thing, and it should be avoided (apathy). Passions inhibit by fixating attention on a single inclination and blinding one to all other inclinations. They are consistent with means–end reasoning, and even with reasoning about subordinate ends, but, in the case of a passion, all reasoning takes place in the context of one’s overriding passion. Thus passions are a sort of extreme case of radical evil, where one’s fundamental maxim, rather than merely subordinating morality to happiness, subordinates both morality and happiness to the end of a particular inclination. They are properly evil and are “fixed” in the same way that evil itself is.