Kant's character-based account of moral weakness and strength

Carl Hildebrand

Abstract
The standard account of Kantian moral weakness fails to provide a psychologically realistic account of moral improvement. It assumes that moral strength is simply a matter of volitional resolve and weakness is a lack of resolve. This leaves the path to moral improvement unclear. In this paper, I reconstruct an alternative character-based account of Kantian moral weakness and strength. On this account, moral strength is the possession of sympathy and self-knowledge, key practical-epistemic virtues from Kant's *Doctrine of Virtue*, and moral weakness is a lack of these virtues. This identifies moral strength with a high degree of development, integrity, or fitness in one's character, and not merely an ability to somehow try harder. It also resolves an exegetical puzzle concerning the change of heart in Kant's *Religion Within the Boundaries of Mere Reason*.

Keywords
Kant; moral character; virtue; weakness; akrasia; strength; moral improvement

1. Introduction

The standard account of moral weakness in Kant locates that weakness in an insufficient resolve to carry out the commands of the moral law. Whether this is understood in terms of an unstable higher-order commitment to the moral law, as Adam Cureton, Marijana Vujošević, and Patrick Frierson have argued, or a conflict between values and motivations, as Robert Johnson has argued, each of these approaches defines moral strength as a matter of resolve in fulfilling maxims of duty. And conversely, each defines moral weakness as a lack of such resolve. One of the challenges of understanding moral weakness in this way is that it leaves the path to moral strength unclear. If moral strength is akin to gritting one's teeth and just doing the right thing, it is not abundantly clear how one should begin to do this better. An alternative, comparatively neglected approach is to locate Kantian moral weakness in a defective state of character. Thomas Hill, in contrast with the standard account, has suggested that weakness of will should be analyzed as a feature of moral character in the broader sense, rather than a discrete act of the will

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

to be analyzed afresh in every situation (Hill, 1986, p. 94-95).[1] He applies this suggestion to a reading of Kant, arguing that Kant understands weakness of will not as a lack of power to do something, but as a vague resolution to be moral. This vagueness blurs the content of what morality requires while it weakens one's resolve (Hill, 2008, p. 223). Patrick Frierson has argued that Hill's account does not deliver on its claims—it does not offer a character-based account of moral weakness. In his words, Hill does not "explain an inner mechanism at work, effective in some persons (the virtuous) but not in others (the morally weak)", (Frierson, 2015, p. 236). While a more charitable reading of Hill's account is possible, Frierson is correct that more could be said to explain the psychological gap between moral weakness and strength on Hill's account of Kant.

In this paper, I will extend Hill's line of thought by developing a Kantian account of moral weakness based on defective character, in contrast to the standard account based on volitional resolve. Kant's most focussed discussion of moral weakness occurs in Part One of *Religion Within the Boundaries of Mere Reason* (*Religion*), which is the primary focus of this paper. Here, he discusses the several topics of moral frailty, character, and the change of heart.[2] It is reasonable to believe that each of these phenomena is related, since each in some way addresses the nature of an agent's commitment to the moral law. The account of moral weakness I develop will draw from Kant's discussion of these topics in the *Religion*. Though the term 'weakness of will' is sometimes used in the literature, Kant has less to say about the weakness of will as such, so while the term may occasionally crop up, I will develop his account of moral weakness more generally. This weakness is apparent in a puzzle that arises from his discussion of the change of heart in the *Religion*. I argue this puzzle may be resolved and an account of moral strength (and implicit weakness) reconstructed based on Kant's notion of character, specifically sympathy and self-knowledge, from the *Doctrine of Virtue.* I will limit my account to these two works, the *Religion* and the *Doctrine of Virtue*.[3]

It is helpful to briefly describe the related phenomena of moral frailty, character, and the change of heart in the *Religion*, and the puzzle that arises there, before stating my argument concerning moral weakness in more detail. Moral frailty is the first (i.e., least pernicious) grade of propensity to evil in human nature, described as "the general weakness of the human heart in

[1] In one sense, the standard account could be taken to define moral weakness as a character defect, since character is in part defined by Kant as a commitment to the moral law; hence insufficient commitment implies defective character. See, for example, the *Anthropology*: "character requires maxims that proceed from reason and morally-practical principles" (Kant, 7:293), or the opening to *Groundwork* I (Kant, 4:393-394). I will discuss this in more detail below. In another sense, however, it does not explain it in terms widely recognized as features of character beyond the confines of Kant scholarship.

[2] The related topic of volitional resolve is not as explicit here as elsewhere in Kant's corpus, though it is implied in his discussion of frailty and features prominently in commentators' discussion of frailty (for two examples discussed here, see Cureton, 2016, or Vujošević, 2019). This is more explicit, for example, in the *Doctrine of Virtue* when he says that it is "the strength of one's resolution, in the first place, that is properly called *virtue*" (Kant, 6:390).

[3] It might be argued that Kant's position on matters of virtue, character, weakness, and strength differs between these two works. While he may at times use the same or similar language in different ways, it is unlikely that his deeper views differ importantly between these texts, as they were published only five years apart from one another, at a similar stage in his career and thought.

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

complying with the adopted maxims" (Kant, 6:29). The change of heart is discussed in Part One Section IV, Concerning the Origin of Evil in Human Nature, in the General Remark, which discusses the restoration of human nature from evil to good. It involves a change in one's fundamental maxim (*Gesinnung*) from self-love to the moral law. It is also described as the acquisition of virtue in one's intelligible character, beyond virtue of merely empirical character (Kant, 6:47). The change of heart is therefore a change of character and Kant believes it carries implications for the evaluation of an agent's character. However, it is less clear what those implications are, or how we should understand the nature of intelligible (noumenal) and empirical character. It is useful to look at his description of the change of heart and empirical and intelligible character in the text. He begins by describing empirical character.

> When the firm resolve to comply with one's duty has become a habit, it is called a *virtue* also in a legal sense, in its *empirical character* (*virtus phaenomenon*). Virtue here has the abiding maxim of *lawful* actions, no matter whence one draws the incentives that the power of choice needs for such actions. [….] But not the slightest *change of heart* is necessary for this; only a change of *mores*. A human being here considers himself virtuous whenever he feels himself stable in his maxims of observance to duty—though not by the supreme ground of all maxims (Kant, 6:47).

To change the supreme ground of all of one's maxims is to undergo a change of heart.

> However, that a human being should become not merely *legally* good, but *morally* good (pleasing to God) i.e. virtuous according to the intelligible character [of virtue] (*virtus noumenon*) […] cannot be effected through gradual *reform* but must rather be effected through a *revolution* in the disposition [*Gesinnung*] of the human being (a transition to the maxim of holiness of disposition) (Kant, 6:47).

This revolution of disposition is the change of heart. Here, Kant only briefly suggests an answer to the question of how we should evaluate an agent who has undergone the change of heart, whose noumenal character has changed for the good but whose empirical character does not show obvious improvement: he states that while a "revolution is necessary in the mode of thought […] gradual reformation [must be possible] in the mode of sense" (Kant, 6:47-8).[4] Here is a puzzle: what does this gradual reformation look like; how does one make moral progress? After the change of heart an agent may have a good disposition (*Gesinnung*), being fundamentally motivated by the maxim of duty, but unable to consistently translate this into action that honors the worth of others. This might appear to make strange use of the word character as a term of praise, since someone with 'a good character' is commonly thought of as someone who has more than just good intentions.[5]

I will rationally reconstruct an account of virtuous empirical character based on the account of virtue in the *Doctrine of Virtue*, in particular the virtues of sympathy and self-

[4] Kant's works will be cited by volume and page number of the standard Akademie edition (Berlin, 1900).

[5] This is a gloss on common usage. It is reasonable to assume that the standard for an ascription of virtuous moral character is high, as classically demonstrated, for example, in Aristotle's *phronimos* (Aristotle, 1999, 1141a).

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

knowledge—wide duties of love cashed out as moral aptitudes, or proficiencies the virtuous must cultivate. The acquisition of these virtues may be understood to fill the gap between virtue of intellectual and empirical character. Based on these texts, Kantian virtuous character may be summarized as involving two things: (a) a commitment to the moral law as one's fundamental maxim and (b) well-developed aptitudes of sympathy and self-knowledge. In line with this, I want to suggest that moral frailty may be better understood along the lines of character than volitional resolve. Importantly, Kantian sympathy and self-knowledge are acquired through time. Once one commits to following the moral law, one commits to acquiring these virtues and therefore improving morally.[6] Whereas it is unclear how one might become better at simply willing to do the good as the standard account suggests, it is more obvious how one may grow in sympathy and self-knowledge. These virtues illuminate the path to moral improvement for Kant, identifying moral strength with a high degree of development, integrity, or fitness in one's character. Conversely, I argue contrary to the standard account that Kantian moral weakness may be understood as a deficient development of one's empirical character, also understood in terms of sympathy and self-knowledge. This does not have to contradict the standard approach to moral weakness—it should not be surprising if there is more than one way to be weak—but it does provide a richer explanation in terms of moral psychology. It also provides a clearer path to moral improvement and solves an interesting puzzle concerning the change of heart.

2. The standard approach

The standard approach to Kantian moral weakness locates that weakness in an insufficient resolve to carry out the commands of the moral law. This might seem to follow naturally from Kant's distinction between three grades in the propensity to evil.

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

Here, moral weakness (frailty) appears to be a straightforward failure to comply with moral maxims. Prima facie, that failure might follow from two things: one, failure to act on one's chosen moral maxim (i.e., acting against one's own better judgment), or two, failure to consistently will the moral maxim in the first place (i.e., acting from shifting maxims—with an unstable commitment). The challenge with the first possibility is that it appears to run contrary to

[6] I distinguish between holiness and what might be called full virtue in Kant. I take holiness to be an attribute of a will that is "of itself necessarily in accord with the law" (Kant, 4:14), or a divine will (Kant, 5:82). A finite rational being—one that is both intelligible and empirical—cannot attain holiness, even if it has undergone a change of heart and made a "transition to the *maxim* of holiness of disposition" (Kant, 6:47, emphasis mine). To have what might be called full Kantian virtue, I will argue, is to have virtue in both empirical and intelligible character. The former entails the above transition to the maxim of holiness; the latter entails the possession of the moral aptitudes of sympathy and self-knowledge in a high degree of development. Full virtue in this sense does not entail holiness, perfection, or complete goodness, and can therefore be attributed to a finite rational being.

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

the incorporation thesis. Drawn from the *Religion* and defined by Henry Allison, the incorporation thesis holds that "an inclination or desire does not *of itself* become a reason for acting" until an agent has incorporated it into her maxim in an act of spontaneity (Allison, 1990, p. 40).[7] The challenge this poses is that it appears to rule out the possibility that an agent may act against their own better judgment. Since one cannot act without incorporating an incentive into their maxim, it appears they simply change their mind. So, if moral weakness involves acting against one's own better judgment, as it is standardly taken to do, the incorporation thesis appears to rule out weakness of will.

There are ways to address this challenge. One may, for example, modify the incorporation thesis and argue that it does not rule out weakness of will, as Robert Johnson does (see below). However, proponents of the standard account typically address this by identifying volitional resolution with the stability of the higher-order maxims an agent incorporates.[8] For example, Adam Cureton argues that strength of will should be understood as the strength of one's basic commitment to the moral law, where strength is understood as stability. He suggests that

> someone's will is stable or strong if she tends not to alter her basic commitment too readily and she tends to revert back to it were it to change, while a person's will is unstable or weak if she tends to alter her basic commitment too readily and tends not to revert back to it were it to change (Cureton, 2016, p. 71).

While this makes sense on logical grounds, it is less convincing as an interpretation of Kant. There is little, if any, textual evidence to indicate that our basic disposition can change more than once or twice. In any case, it is rare for these changes to occur and more common to simply act in a manner inconsistent with our basic disposition.[9] If it were nevertheless granted that he identifies a sufficient condition for moral weakness on the Kantian picture, he does not identify a necessary condition. In other words, Cureton identifies one form of moral weakness in Kant while there remain other possibilities. A further challenge to this account is that it leaves the path to moral improvement unclear.

> Because we can never know with much certainty whether we or anyone else has a good and stable will, we can never know whether our duty of moral self-improvement is satisfied. Our best option, in light of this ignorance, is to continue striving to adopt and maintain a good will and hope that we will be successful in doing so (Cureton, 2016, p. 72).

It remains unclear what constitutes this striving beyond simply trying harder to do the right thing.

---

[7] This is based on a passage from Kant, 6:27.

[8] Iain Morrison argues, on the contrary, that Kant contains an account of "non-justifying maxims" which explains moral weakness in a way that bypasses the challenge from the incorporation thesis (Morrison, 2005, p. 74-75). His argument addresses non-moral maxims only, so applies only to the "non-moral sphere" (Morrison, 2005, p. 74-75). Consequently, I will not address his account here as my primary concern is with *moral* weakness.

[9] I thank Stephen Palmquist for drawing my attention to this point in personal correspondence.

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

Similar to the description of moral weakness in terms of an unstable will, Marijana Vujošević argues that the morally weak agent "gives priority to sensible incentives in his maxims, but he does so by merely failing to renew his commitment to the moral law by reassessing his incentives in new situations" (Vujošević, 2019, p. 48). She explains this as a problem in motivation: the weak agent "fails to incorporate the law as the self-sufficient incentive in his supreme or underlying maxim, and his heart can therefore be characterized as morally evil" (Vujošević, 2019, p. 47). Neglect to renew commitment to the moral law accompanied by a readiness to reassess one's incentives (especially amidst opportunities to advance self-interest) is one way to exemplify weakness. This is a variation of the standard account because it identifies moral weakness with a lack of resolution and explains that lack of resolution as an unstable commitment to the moral law. Vujošević holds that the weak agent's heart is morally evil, which she takes to follow from Kant's rigorism, the idea that one's fundamental maxim must be either good or evil, the moral law or self-love, exclusively (Vujošević, 2019, p. 47).

As per Kant, this is one instantiation of the *propensity* to evil. However, a weak agent is typically one who in some sense still wants to do good. So, while in a sense Vujošević might be right that an agent's motivation is evil at that moment they have *succumbed* to weakness, this does not describe their motivation qua weak agent generally. To do so would require a more general assessment of that agent's motivations through time along with a distinction between their motivation and inclinations. Vujošević is therefore too quick to identify the weak agent as evil. A weak agent must in some sense be conflicted. This is indicated in Kant's claim that "the frailty (*fragilitas*) of human nature is expressed even in the complaint of an Apostle: 'What I would, that I do not!'" (Kant, 6:29). One way this conflict can be accommodated is by examining an agent's motivations over time and according to their considered judgment. This is what Richard Holton does, whose idea of weakness of will is reflected in the standard approach (often implicitly), including Vujošević's account (implicitly) (Holton, 1999, p. 247-248). It is therefore incorrect to identify the weak agent as evil in this way. While Vujošević is right that rigorism rules out a mixture of motives at the level of fundamental motivation, there are other regions of an agent's psychology where this conflict may be located. For example, between an agent's (freely chosen) fundamental intention and their naturally determined inclinations, or, as I will argue below, between their fundamental intention and the understanding requisite to successfully follow through on it. Further, this account too does not provide direction for how to improve and grow in moral strength, aside from simply advising a higher degree of self-control or commitment to one's moral maxims.

Patrick Frierson explicitly draws from Holton in his interpretation of moral weakness. He argues that Holton's policy intentions (from Michael Bratman) are roughly equivalent to Kantian maxims, pointing out that for Kant, to have character requires that one have stable and consistent life-guiding maxims. A failure to have these maxims is, in one sense, a failure to have any character at all: "in order to be *a* person, one must have a character—and hence principles—that

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

are at least relatively stable and consistent" (Frierson, 2014, p. 243).[10] An agent who is too quick to revise these principles—the Kantian analog of policy intentions—compromises their character and is therefore an agent who is weak. Such an agent has allowed their higher faculties (of desire and cognition) to be manipulated by their lower faculties (inclinations). While they make choices of which they are aware and for which they are responsible, those choices do not reflect their purported life-guiding maxims. In this sense, they lack a character (Frierson, 2014, p. 247-248). While this identifies moral weakness with a defect of character, it defines character exclusively in terms of one's higher-order commitment, or resolve to stick to one's principles. This is one feature of Kant's understanding of character but fails to capture broader features of character developed in other texts of the same period, notably the *Doctrine of Virtue*. Passing over these features further neglects an opportunity to illuminate the path to moral improvement.

Robert Johnson similarly emphasizes the role of character in his account of moral weakness, seeming to depart from the standard approach. He argues for a broader view of the incorporation thesis on which it "concerns not merely the incorporation of various incentives into our *motives*, but also into our *values*" (Johnson, 1998, p. 358). This is based on a threefold distinction between motivations for particular actions, valuations of actions (on which an agent may fail to act), and objective laws valid for any agent (namely, the moral law) (Johnson, 1998, p. 357). It allows that an agent may incorporate moral principles into their valuation of actions while in some cases failing to be motivated by those principles. Such an agent may act against their own better judgment without contradicting the incorporation thesis because they have incorporated the *valuation* of a particular action but not the wayward *motivation* on which they act. Johnson argues that such a motivation remains unincorporated. On this account, the conflict within a weak agent lies between "the *values* enshrined in an agent's character and what *motivates* her", rather than maxims issued from their higher and lower faculties, respectively (Johnson, 1998, p. 361).

Johnson suggests that "a *Gesinnung* in which there is a proper order of incentives is quite compatible with a lack of (empirical) virtue" because one may value the right things without being motivated by them (Johnson, 1998, p. 359).[11] To have empirical virtue then is a matter of being motivated by the good values one already has (presuming one already has those values). In this way, empirical virtue once again amounts to volitional resolution. Functionally, Johnson's character-based account becomes another form of the standard account. The difference is that instead of defining weakness in terms of unstable principles or "erratic behaviour," Johnson defines it in terms of motivations that are out of sync with one's values.[12] If empirical character is about following through on one's moral commitments, this does explain why the morally

[10] Frierson points out that Kant uses the term 'character' in two senses: in one sense, character is something that every agent with a will (higher faculty of desire) has, in another sense, it is a rare accomplishment of firm and stable life-guiding maxims (Frierson, 2014, p. 246). Moral weakness is defined by lack of character in the second sense.

[11] He continues: a "person's genuinely pure *Gesinnung* may not, in other words, have a sufficient influence on what motivates her in particular situations, though were she fully rational, her own good will would be 'irresistible'" (Johnson, 1998, p. 359).

[12] See page 360, where he argues that weakness of will must be distinct from "erratic behaviour" (Johnson, 1998).

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

weak, who have the right commitments but lack empirical character, are better than the vicious (Johnson, 1998, p. 351-352). This has the advantage of avoiding the problem around Vujošević's identification of weakness with an evil intention (*Gesinnung*).[13] However, it does not offer a clear path to moral progress because it identifies strength once again merely with volitional resolve.

3. The role of sympathy and self-knowledge in completing character

Empirical character is more than a matter of volitional resolve, or so I will argue. Kant draws a sharp distinction between intelligible and empirical character when he discusses the change of heart in the *Religion*.[14] I suggest that Kantian moral weakness and strength is better understood as that which completes the gap between virtue in intelligible character and virtue in empirical character. Once an agent has undergone the instantaneous change of heart and their intelligible character becomes good, they must labour through time for their empirical character to be good as well. The problem is that Kant provides only scarce and suggestive remarks as to how this should be done. It is also unclear how we are to evaluate an agent in this interim position. This portion of Kant's text therefore presents us with a puzzle. In this section, I will reconstruct a solution to this puzzle that lays the ground for a character-based account of moral weakness and strength.

But first I will describe the puzzle. Kant says that to become morally good a person must undergo a "revolution" in their fundamental disposition (*Gesinnung*).[15]

> So long as the foundation of the maxims of the human heart remains impure, [becoming morally good] cannot be effected through gradual *reform* but must rather be effected through a *revolution* in the disposition of a human being (a transition to the maxim of holiness of disposition). And so a 'new man' can come about only through a kind of rebirth, as it were a new creation […] and a change of heart. But if a human being is corrupt in the very ground of his maxims, how can he possibly bring about this revolution by his own forces and become a good human being on his own? Yet duty commands that he be good, and duty commands nothing but what we can do. The only way to reconcile this is by saying that a revolution is necessary in the mode of thought but a gradual reformation in the mode of sense (which places obstacles in the way of the former), and [that both] must therefore also be possible to the human being (Kant, 6:47).

[13] Vujošević would be correct if, instead, she took frailty to be evidence of a *propensity* to evil (Kant, 6:29).

[14] The idea of the change of heart is odd to many readers of Kant, independent of its roots in Christian theology. Quassim Cassam has argued that something like this idea has merit on independent philosophical grounds: see chapter 8 of Cassam, 2019.

[15] For an alternative translation of Kant's *Gesinnung*, see Palmquist, 2015.

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

He is clear that gradual change for the better in empirical character is possible, but the details as to how this change takes place are unclear. He goes on to say:

> From this it follows that a human being's moral education must begin, not with an improvement of mores (*Sitten*), but with the transformation of his attitude of mind and the establishment of a character, although it is customary to proceed otherwise and to fight vices individually, while leaving their universal root undisturbed (Kant, 6:48).

This supports the distinction between empirical character and intelligible character, each of which possesses its own form of virtue. Contrary to Frierson's account which identifies virtuous character with stable and consistent (moral) maxims, Kant here provides us with two ways for an agent to possess virtuous character, each of which appears to be necessary for virtuous character in the full sense. To have virtue of intelligible character is to have virtue in the primary sense, whereas to have virtue of empirical character is to have virtue in a secondary sense. To have virtue in both intelligible and empirical character is to have virtue in the full sense.

The virtue of empirical character "has the abiding maxim of *lawful* actions, no matter whence one draws the incentives that the power of choice needs for such actions" (Kant, 6:47). It is acquired incrementally through a gradual process of habituation. It results in a human agent having passed from "a propensity to vice to its opposite" and having developed a sense of stability in the maintenance of dutiful actions (Kant, 6:47). However, it does not imply that one's foundational maxim is moral: as Kant says, it does not imply a change of heart from bad to good but only a change of mores (Kant, 6:47). While a liar might begin to tell the truth for the sake of reputation, it does not make sense to say that such a person has *virtuous* character, even if they tell the truth consistently (Kant, 6:47). One's external actions may conform to duty while one's fundamental maxim or disposition remains evil (self-love remains the condition for 'good' behavior). Virtue in one's empirical character does not entail virtue in one's intelligible character; this would require that one's fundamental maxim be moral (the moral law).[16]

It is easy enough to understand how one may have virtue in the external or empirical sense and yet lack virtue in the internal or intelligible sense (i.e., not be motivated to do the virtuous thing because it is virtuous). Examples like Kant's self-interested honest man (Kant, 6:47) are familiar to us. Yet the dual nature of character found in this text allows the order of virtue to go in the other direction too and this is more puzzling. After undergoing the instantaneous change of heart, the transformation of empirical character takes time: "a revolution is necessary in the mode of thought but a gradual reformation in the mode of sense […] he is to this extent, by principle and attitude of mind, a subject receptive to the good; but he is a good

[16] Kant holds that such a change in intelligible *Charakter*, and hence in one's fundamental maxim, cannot happen gradually as it may happen in one's empirical *Charakter*. This has to do with his idea of rigorism, which forbids that maxims be mixed, so requires that one's fundamental maxim be either good or evil (where the 'or' is taken in the exclusive sense). So, there is no way for such a change to happen except suddenly, as if by a kind of "revolution" or "single and unalterable decision" (Kant, 6:47-48). The result of this change of heart is that one "should become not merely *legally* good, but *morally* good (pleasing to God) i.e., virtuous according to the intelligible *Charakter*" where "for God, this is the same as actually being a good human being" (Kant, 6:48).

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

human being only in incessant laboring and becoming" (Kant, 6:48). What is the object of this laboring and becoming? How is it that one who has undergone the change of heart and whose fundamental motivation is therefore good can still not get things right, so must engage in this process of continual self-improvement? Or how exactly can a person with a good will have what appears to be bad character?[17]

Because it operates at the level of intelligible character, the change of heart involves an agent's fundamental maxim, which defines their character in the deepest sense. Since it involves a maxim and so an intention, the change of heart must also involve a capacity for choice.[18] An agent's good or evil character must be "an effect of his free power of choice," otherwise his character could not be imputed to him (Kant, 6:64).[19] The change of heart therefore involves a fundamental commitment to prioritize the good maxim of the moral law over the maxim of self-love.[20] It is a choice for the good, a choice to make the moral law the condition of one's action in the sense that all other reasons for action, including self-love, are subordinate to it (Kant, 6:36). When it is moral, Kant describes this fundamental maxim as a good character [*Charakter*] acquired by choice.[21] This fundamental maxim, also understood as intelligible character, represents the first, primary element in Kant's conception of moral character. It can be evaluated as good or bad, requires choice on the part of the agent, and is imputable to the agent.

The secondary elements in Kant's conception of moral character may be understood as features of empirical character. Once the change of heart has occurred, that one's fundamental maxim has been converted to the good (meaning that one prioritizes the moral law over self-love) this entails that the individual must embark on a journey of moral self-formation to become thoroughly good. To become thoroughly good is to rightly prioritize one's incentives, enabling the moral law to control self-love. The change of heart makes one "a subject receptive to the

[17] One might object that this is simply to say that will is distinct from character, which is just Kant's position contrary to Aristotle, say, whose understanding of will, choice, and character hang together quite differently. However, we should ask why Kant uses the term character (*Charakter*) in both instances, in the empirical and intelligible sense. And even if we accepted this, replacing the term character with will at the noumenal level, there would remain something odd in our understanding of character at the empirical level, since to have good character is usually to be a good person in a sufficiently thorough sense, including one's will. Conversely, we do not typically say that people with bad character have a good will.

[18] Taking intention to be implied in a "subjective principle of volition" as per Kant's definition of a maxim in Kant, 4:401.

[19] Here I take Kant to be referring to character at this juncture, following George di Giovanni, who includes the term in square brackets in his translation of the text: "These two [characters] must be an effect of his free power of choice".

[20] See his discussion of the disjunctive proposition and 'rigorism' in Kant, 6:22 and mentioned in note 16 above. Whether one is evil or good is then a matter of which maxim, self-love or the moral law, is made the condition of the other, as he says in Kant, 6:36.

[21] "But now this is possible only because the free power of choice incorporates moral feeling into its maxim: so a power of choice so constituted is a good character [*Charakter*], and this character [*Charakter*], as in general every character of the free power of choice, is something that can only be acquired" (Kant, 6:27). Elsewhere he describes this "first ground", whether good or evil, as that by which the human being expresses the character (*Charakter*) of their species (Kant, 6:21).

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

good; but he is a good human being only in incessant laboring and becoming" (Kant, 6:48). To become a good human being (*ein guter Mensch*) involves continual work through time, indicating that one is not yet an entirely good human being once one has undergone the change of heart. Good character is therefore not complete until features beyond one's fundamental motivation (*Gesinnung*) have been shaped to align with that motivation, to enact and enhance its activity. Kant describes this formation in terms of moral education, which must begin not with an individual's improvement in mores (*Sitten*), but with the "transformation of his attitude of mind [*Denkungsart*] and the establishment of a character [*Charakter*]" (Kant, 6:48).

He goes on to say that this predisposition "gradually becomes an attitude of mind [*Denkungsart*]", though he does not specify here how this happens (Kant, 6:48). The concept of moral aptitude [*moralische Fertigkeit*] from the *Doctrine of Virtue* provides the clearest picture of what a moral *Denkungsart* would look like. Kant defines moral aptitude as "a facility for acting and a subjective perfection of *choice*" in which one "determines oneself to act through the thought of the law" (Kant, 6:407). He contrasts it with habit, stipulating that only the former may count as virtuous.

> An *aptitude* [*Fertigkeit*] (*habitus*) is a facility in acting and a subjective perfection of choice. – But not every such *facility* is a *free* aptitude (*habitus libertatis*); for if it is a *habit* [*Angewohnheit*] (*assuetudo*), that is, a uniformity in action that has become a *necessity* through frequent repetition, it is not one that proceeds from freedom, and therefore not a moral aptitude [*moralische Fertigkeit*]. [….] Only such an aptitude can be counted as virtue (Kant, 6:407).

Virtue, according to Kant, therefore, takes on a predominantly cognitive character. It is more about having the right thoughts and disposition of mind (a moral aptitude or *moralische Fertigkeit*), rather than only a behavioral disposition to perform certain actions (however beneficial the outcome of those actions may be) (Kant, 7:149).[22]

The duties of love to other human beings, also spelled out in the *Doctrine of Virtue* to substantiate this idea of virtue, add content to this idea of moral aptitude. Love is understood as the "maxim of *benevolence* (practical love), which results in beneficence" or "making the well-being and happiness of others my *end*" (Kant, 6:450). It has both an internal and external aspect, requiring that an agent be motivated by the happiness of others and act properly based on that motivation. Action that benefits others without intention to do so is mere habit, as described above, while an intention to benefit others without action to do so is incomplete. The duties of love include beneficence, gratitude, and sympathy.

[22] Habit is an unthinking pattern of action that results from frequent repetition of that action: it does not involve choice on the part of the agent. Because virtue requires choice, habit cannot count for virtue. He also suggests that habits of this kind typically restrict one's freedom, saying in the *Anthropology* that "habit [*Angewohnheit*] deprives even good actions of their moral worth because it impairs the freedom of the mind" (Kant, 7:149).

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

The duty of sympathy is particularly informative.[23] Kant acknowledges that it is natural to share sympathetically in the feelings of others, experiencing joy in their pleasure and sadness in their displeasure. He argues that we have a duty to use these sympathetic feelings "as a means of promoting active and rational benevolence" (Kant, 6:452). This duty remains a conditional duty, since it rests on the empirical fact that we are beings who experience these feelings and are therefore able to use them. For example, he mentions elsewhere, in the *Anthropology Friedländer* lectures, that if we had a greater degree of reason, we would not need sympathy, because we could see directly what leads to the well-being of others.[24] This suggests that the primary purpose of sympathy is to help us *understand* the experience of others, and these feelings provide insight into that. So, since we experience these feelings and can leverage them for moral purposes, we should do so, using them insofar as they help us to better fulfill our duty. Kant takes a hard line on this, maintaining that if we are unable to help another who is suffering, we ought not to indulge sympathetic feeling for them, since it leads to more suffering overall:

> when another suffers and, although I cannot help him, I let myself be infected by his pain (through my imagination), then two of us suffer, though the trouble really (in nature) affects only *one*. But there cannot possibly be a duty to increase the ills in the world (Kant, 6:457).[25]

We may infer from all this that what is unconditional in Kantian sympathy is that it entails understanding and performance of moral action, while the fact about whether it involves feeling is conditional.

Sympathy, including sympathetic feeling, aids us in the performance of moral action primarily by providing us with information about the well-being of others—what contributes to their joy and suffering—so that we may act to remediate their suffering and increase their joy. This is what he means by sympathizing actively in their fate.

> But while it is not in itself a duty to share the sufferings (as well as the joys) of others, it is a duty to sympathize actively in their fate; and to this end it is therefore an indirect duty to cultivate the compassionate natural (aesthetic) feelings in us, and to make use of them as so many means to sympathy based on moral principles and feeling appropriate to them (Kant, 6:457).

He continues:

> It is therefore a duty not to avoid the places where the poor who lack the most basic necessities are to be found but rather to seek them out, and not to shun sickrooms or debtors' prisons and so forth in order to avoid sharing painful feelings one may not be able to resist. For this is still one of the impulses that

[23] The duties of beneficence and gratitude appear far less interesting from a Kantian point of view, as beneficence explains what is already implicit in the more general duty to love, while the duty of gratitude amounts to a duty to respect others, including those who give good things.

[24] "If we were beings who had a greater degree of reason, then we would not need any sympathy, for we could have insight into the other's well-being or misfortune from the principles. Sympathy is therefore only a means of supplementing the lack in principles" (Kant, 25:611).

[25] Kant repeats the same example elsewhere (e.g., in the *Anthropology Friedländer* lectures; Kant, 25:611-612).

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

> nature has implanted in us to do what the representation of duty alone might not accomplish (Kant, 6:457).

Kantian sympathy therefore performs a primarily epistemic function by providing cognitive content about the world. It informs us about the status of others' well-being and features of the world that contribute to or detract from it. To be sympathetic in the Kantian sense is to have an understanding of these features of the world, alongside a motivation to act for the well-being of others.[26] This sympathetic proficiency substantiates Kant's idea of moral aptitude earlier in the *Doctrine of Virtue*, since the agent with well-developed sympathy has a facility for moral action which follows from an advanced cultivation (perfection) of their subjective nature. The acquisition of this proficiency is something that occurs through time.

Melissa Merritt suggests a different interpretation, according to which "[w]hat gets cultivated seems to be Humean-style sympathy, a natural propensity for the communication of feeling" (Merritt, 2018, p. 195). She correctly observes that such sympathy is grounded in practical reason and as a means for practical benevolence. She provides an account according to which "for the natural inclination to sympathy to become *skilled*, it must be that it becomes more responsive in concrete and finely grained ways" and one way this may happen "is through our close relationships with *particular* other people—people whom we know well, and in whose well-being and happiness we naturally take a visceral interest" (Merritt, 2018, p. 196). This is insightful. Sympathy is acquired and cultivated in the context of personal relationships, often close ones, in which we acquire an interest in and understanding of the well-being and happiness of others. However, while this sympathy may at times involve visceral feelings, it need not necessarily: as I have argued, the involvement of feeling in Kantian sympathy is conditional. What is unconditional is that it involves understanding and performance of moral action. Kant is endorsing a form of what nowadays might be called cognitive empathy, something more akin to social intelligence than something based on feeling.[27] It is therefore incorrect that Kantian sympathy necessarily entails the communication of feeling, as it does on Hume's account.[28]

Melissa Seymour Fahmy goes further than Merritt, arguing that Kant's duty of sympathy is a direct duty to produce and express sympathetic feelings, and to engage others "affectively" (Seymour Fahmy, 2009, p. 47). Nancy Sherman similarly argues that we are obligated to

---

[26] Kant emphasizes that sympathetic feeling proper must be free, based on an agent's choice, in the following passage: "Humanity can be located either in the capacity and the will to share in others' feelings (*humanitas practica*) or merely in the receptivity, given by nature itself, to the feeling of joy and sadness in common with others (*humanitas aesthetica*). The first is free, and is therefore called sympathetic (*communion sentiendi liberalis*); it is based on practical reason. The second is unfree (*communion sentiendi illiberalis, servilis*); it can be called communicable (since it is like receptivity to warmth or contagious diseases), and also compassion, since it spreads naturally among human beings living near one another. There is obligation only to the first" (Kant, 6:456-457).

[27] See, for example, Paul Bloom's discussion of the difference between cognitive and affective empathy in chapter 1 of Bloom, 2016.

[28] See, for example, the explanation of sympathy offered in *A Treatise of Human Nature,* SB-316-321 (Selby-Bigge pagination), in which sympathy is defined as the communication of sentiment. "No quality of human nature is more remarkable, both in itself and in its consequences, than that propensity we have to sympathize with others, and to receive by communication their inclinations and sentiments, however different from, or even contrary to our own."

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

manifest sympathetic emotions in beneficent actions (Sherman, 2014, p. 22). However, this goes beyond what Kant is concerned with in the *Doctrine of Virtue*. It is not clear in what he says that we have a duty to express *feeling*; further, it seems there could be no such duty if sympathy is "only a conditional duty"—that is, conditional on our ability and willingness to act for the sake of the other (Kant, 6:456). He says that "while it is not in itself a duty to share the sufferings (as well as the joys) of others, it is a duty to sympathize actively in their fate" (Kant, 6:457). This is also implied in the hard line taken in the Stoic example above.

Beyond sympathy, moral cognition of oneself further substantiates the idea of moral aptitude in the *Doctrine of Virtue*. Kant describes it as a command and the first among all duties to oneself, to know yourself "in terms of your moral perfection in relation to your duty" and "to know your heart – whether it is good or evil, whether the source of your actions is pure or impure" (Kant, 6:441). He goes so far as to say at one point that it is the beginning of all human wisdom.[29] While this moral cognition implies respect for oneself, it also serves to "counteract that *egotistical* self-esteem which takes mere wishes—wishes that, however ardent, always remain empty of deeds—for proof of a good heart" (Kant, 6:441). To be virtuous in the Kantian sense requires that one know oneself, including one's own maxims, motivations, and patterns of thought, so that any features of oneself conflicting with the moral law (e.g., by serving self-interest at the expense of others) might be expunged from one's character. Though Kant here uses the language of knowledge (*Erkenntnis*) and moral cognition (*moralische Selbsterkenntnis*), the knowledge of which he speaks cannot strictly speaking be complete. Paradoxically—and insightfully—this kind of Kantian self-knowledge entails an awareness of one's own ignorance of oneself. Though complete knowledge cannot be obtained, one should never cease from scrutinizing, fathoming, and *seeking* "to penetrate into the depths […] of one's heart" (Kant, 6:441). If one does not seek self-knowledge, it remains possible and perhaps likely that one will overlook opportunities for beneficent action, among other things.

Lack of self-knowledge also leaves one vulnerable to self-deception, for example, in cases in which one believes oneself to be acting for the sake of another when one would not perform the action if it did not, for example, garner esteem in the eyes of others. Even the virtuous agent is susceptible to this, as it is impossible to eliminate the propensity to evil (Kant, 6:51). They must therefore remain vigilant, continually seeking to understand themselves, their motives, and "the depths (the abyss) of [their] heart which are quite difficult to fathom" (Kant, 6:441).

In this way, both sympathy and self-knowledge account for what it means to have moral aptitude in the Kantian sense. Knowledge of the world, others, and oneself is necessary to properly execute moral action. An agent approximates full virtue in the Kantian sense when they have developed an aptitude for understanding and acquiring this information, alongside a commitment to acting on it. This enables us to say that full virtuous character on the Kantian picture requires properly developed aptitudes of sympathy and self-knowledge. These features

[29] "Moral cognition of oneself which seeks to penetrate into the depths (the abyss) of one's heart which are quite difficult to fathom, is the beginning of all human wisdom" (Kant, 6:441).

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

are typically acquired after a change in one's fundamental motivation and, in any case, commitment to the moral law entails commitment to their acquisition. In this sense, they constitute secondary features of virtuous character in addition to the primary feature of commitment to the moral law. Though they are no less necessary for full Kantian virtue.

4. Explaining moral failure and improvement

Returning to the puzzle in the *Religion*, the acquisition of sympathy and self-knowledge is a project on which one who has undergone the change of heart must embark. These features of an agent's character provide an object of "laboring and becoming" (Kant, 6:48). They explain the gap between noumenal and empirical character, demonstrating how one who has undergone the change of heart may have virtuous character in the noumenal sense yet not in the empirical sense: an agent in this position has not yet fully developed adequate sympathy and self-knowledge.

In the *Religion*, one who has acquired character in the secondary sense is somewhat vaguely described as having a moral attitude of mind, or *Denkungsart*. The term *Denkungsart*, literally 'manner of thinking,' is used several times in this passage and belongs to the same cluster of concepts as aptitude, or *Fertigkeit*, so to substitute one for the other is entirely natural.[30] As I have argued, moral aptitude provides a useful clarification of what a moral *Denkungsart* looks like. We may further distinguish between two types of aptitude, one that is grounded in morality and one that is not. A moral aptitude, like sympathy above, is grounded in (and directed by) the moral law. It does not preclude action for the sake of self-love, so long as that love is rightly ordered, subordinate to the moral law. Not all aptitudes are grounded in morality. The character of Sulla, mentioned in the *Anthropology*, provides an example of an aptitude that is not grounded in morality. Sulla is said to have "firm maxims" and "strength of soul", even though he "arouses disgust" through the violence of his maxims (Kant, 7:293). He exemplifies a form of character insofar as "character signifies that property of the will by which the subject binds himself to definite practical principles that he has prescribed to himself irrevocably by his own reason" (Kant, 7:293). He has great aptitude, perhaps even self-knowledge and sympathy, but this aptitude is grounded, we may assume, in an evil maxim and so is pernicious to morality: his strength of mind and will enables him to perform great violence.

What if someone like Sulla were to undergo a change of heart? It makes sense to hold, as Kant does, that after a change of heart it takes time for an agent with a new commitment to the good to obtain a firm moral aptitude or attitude of mind (*Denkungsart*) consistent with that. It further makes sense to hold that this takes time at least partly because aptitudes or attitudes of mind acquired as the result of a previous prioritization of self-love must be appropriately reworked to adequately serve a new disposition prioritizing the moral law. Aptitudes or attitudes

[30] A notable difference between the two is that *Fertigkeit* entails skill, so is already a term of praise, whereas *Denkungsart* is neutral in this sense (a manner of thinking may be proficient or not). Both refer to features of cognition.

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

of mind that linger in this way may be understood as habits in the unfree sense explained above, resulting from frequent repetition and uniformity in action. They are patterns of thought and so features of an agent's mind, but the agent may not be aware of them and so they are, in a sense, unthinking—that is, unconscious. This much is consistent with, for example, the basic insight behind standard cognitive behavioral therapy. And as any therapist would say, the reworking of these patterns of thought is a process and does not happen in an instant.

Not all these patterns will be transparent to the agent at once, so it will take time for the agent to discover them so that they can change them. The agent will in this way have to engage in a potentially lengthy process of cognitive rehabituation. Throughout this process they might consciously choose to act on the moral law, yet unthinkingly act against it due to the lingering presence of false beliefs that stem from a previous prioritization of self-love. As a variation on Kant's familiar example, we might imagine a wealthy philanthropist who regularly gives money to charities for self-interested reasons. Perhaps these acts give this person a sense of worth precisely because they underscore their superior wealth and social status in contrast to that of their beneficiaries. Perhaps this philanthropist has been an arrogant and unkind person for some time, and their acts of giving are accompanied by condescending and hurtful remarks contrary to the dignity of those who benefit from their gifts. This person might one day realize that these acts of giving are in fact fundamentally motivated by self-love, and they might resolve to act for the sake of the moral law instead, undergoing what Kant would call a change of heart. They might then perform the same acts of giving for the right reasons and yet, without being aware of it, still refer to those who benefit from their donations in a condescending manner contrary to their dignity. They might be so used to speaking this way that they do not recognize it as an instance of moral failure. Perhaps for quite some time they have been surrounded by others who speak this way. They seem to be guilty of a moral failure here, albeit a mitigated one, since it results from a failure in knowledge and not motivation (they have not formed the belief that their remarks are harmful, and they remain unaware that they register morally at all).

On Kant's account, this would be a failure of moral character, because it is a failure of either sympathy or self-knowledge (or both). While this agent has undergone a change of heart, they have not developed the sort of understanding necessary to properly execute moral action. They may have intended good but as far as their action is concerned, mixed with a positive contribution to the well-being of others is a disregard for their dignity. Their motivation, as far as they are aware, remains pure. There is no additional, self-interested condition on which they perform these deeds. So, given that this occurs after the change of heart, it is not a case of impurity in the sense Kant spells out in the *Religion* (their actions prior to the change of heart would be impure).[31] The problem instead is that their lack of sympathy and self-knowledge causes them to do things they do not want to do, namely, to disrespect or harm others. This means it may be read as a case of moral weakness, or frailty: "a general weakness of the human heart in complying with the adopted maxims" (Kant, 6:29). While a change of heart is good—the

[31] The "propensity to adulterate moral incentives with immoral ones" (Kant, 6:29).

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

start of something new—it does not get one all the way to holiness: a person becomes "a good human being only in incessant laboring and becoming" (Kant, 6:48).[32]

It is worth underscoring that this labor is primarily cognitive. If an agent were to examine their own maxims in the way that Kant instructs apprentices in morality to do, and that agent had undergone the change of heart, it would have to be said that they are looking out for habits, beliefs, or attitudes of mind that are the result of a previous disposition of self-love as described, for example, in the case of the wealthy philanthropist above.[33] Though an agent may have undergone the change of heart and made the moral law the supreme ground of their maxims, the propensity to evil which lingers on in even the most virtuous of agents allows maxims of self-love to remain (or enter) at an unconscious level. The virtuous agent may believe they are acting virtuously and desire so to act when, in fact, their action reflects self-love. Though Kant does not directly state here that this attitude of mind is to be identified with sympathy or self-knowledge, in the reconstruction I have provided here, these virtues provide a strong explanation of what it means to have good empirical character.

This leads to an answer to another one of the questions raised at the outset of this paper: how one who has undergone the change of heart and whose fundamental motivation is therefore good can still not get things right, so must engage in a process of continual self-improvement. Answering this involves understanding what virtuous character in the secondary sense does and does not entail. We can understand Kant to be saying that virtuous character in the full sense cannot be approximated without sympathy and self-knowledge. I have argued that, among other things, this entails a cognitive skill or facility in discovering the corrupted state of one's attitude of mind and then, where that corruption lies, correcting it to line up with the good disposition. If this is the case, it is possible for an agent to have a fundamental motivation or intention to enact the moral law then consistently fulfill their moral duties in cases where they know it is their duty to do so, yet due to previous habituation in the contrary direction, remain unaware of other cases in which they ought to do so, or unknowingly act against the moral law, and therefore fail to fulfill their moral duties in those cases. The agent's moral failure in such cases would: first, be an objective moral failure, for example, from the perspective of the moral law and an impartial observer; second, indicate a flawed or incomplete character; and yet, third, may coexist with knowledge of the moral law and an intention to act on it. The final, third point is possible because moral action requires empirical knowledge and proper cognitive habituation, as it were—knowledge of what duty requires in a given situation—in addition to general knowledge of the moral law and the intention to act on it. An agent in such a case of moral failure would not have acquired the moral aptitude, attitude of mind, or knowledge of self, other, and world necessary to enable them to properly execute their fundamentally good intention, and so, for

[32] This example illustrates why Vujošević's position on frailty, which requires an evil heart, is incorrect: the frail agent intends to do good but cannot—they do not intend to do evil.

[33] For Kant's apprentices in morality see Kant, 6:48.

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

reasons of which they are unaware, would fail to properly follow through on that good intention.[34]

There is an illuminating analogy between this sort of case and one interpretation of weakness of will. In both, the agent under consideration appears to have knowledge of what is good and yet fails to act on that knowledge in cases where they ought to do so. Aristotle is held to have in one instance explained weakness of will as what happens when desire interferes with reasoning by preventing relevant information from coming to an agent's attention.[35] If, unlike this, weakness of will were a straightforward matter of seeing two possible courses of action, one motivated by the moral law and the other by desire contrary to it, then choosing to follow the latter, this would indicate that agent to be impure (at best). For it would suggest that one can have virtuous moral character in the primary sense—a good will—and yet that good will can be defeated by contrary desires. Stephen Engstrom, for example, argues that an agent in this state should be understood as having conditioned autonomy—a third category between autonomy and heteronomy. For such an agent the moral principle will succeed in determining their action in some but not all circumstances where contrary desires are present (Engstrom, 1988, p. 452). Such an agent is not unconditionally committed to the moral law, so not autonomous in the strong sense; however, neither are they unconditionally committed to self-love, so they are not heteronomous in the strong sense. Their autonomy is best understood as conditional, or "conditioned" (Engstrom, 1988, p. 448-449). Engstrom argues that conditioned autonomy captures the Kantian idea of impurity, according to which an agent may need "still other incentives besides [the moral law] in order to determine the power of choice for what duty requires" (Engstrom, 1988, p. 452). He further argues that this explains moral improvement through time: an agent advances in virtue as their autonomy becomes less and less conditioned, approximating the ideal of complete autonomy—unconditional commitment to the moral law (Engstrom, 1988, p. 450).

While Engstrom's argument is compelling, it does not address the case of an agent who has undergone a genuine (and we may assume stable) change of heart. An agent in this case remains committed to the moral law even while they have not yet achieved virtuous character. For I am arguing that virtue in the Kantian sense goes beyond autonomy and entails further cultivation of character. This is what I take Kant to be saying in the puzzling passage in the *Religion* (Kant, 6:47-48). The features of character to be cultivated include sympathy and self-

[34] The aptitude described here is necessary for proper execution of moral action regardless of whether Kant is a psychological pluralist or monist; that is, whether he holds that inclination is a motivational capacity distinct from practical reason (as in the former; Schapiro, 2009, p. 233) or understands inclination as an exercise of practical reason (as in the latter; Schapiro, 2009, p. 239). Tamar Schapiro, for example, argues that he is a pluralist, and the logic of the incorporation thesis makes this clear (see Schapiro, 2011, p. 154 for a statement of the argument). In both cases, previous habituation contrary to the moral law may cause one to, in execution, fail to fulfill what the moral law requires despite a fundamental intention to do so.

[35] Aristotle explains weakness of will as resulting from either (i) the non-activation of a minor premise of practical reasoning, or (ii) the possession of a conclusion, but in an off-color way (see analogy to drunk people and students). See commentary in Charles, 1984, p. 127 and VII.3 of Aristotle's *Nicomachean Ethics* for a classic discussion of this.

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

knowledge, as outlined in the *Doctrine of Virtue*. In this way, the autonomous agent may not yet have approximated the virtuous ideal (let alone holiness), in which case they must remain engaged in the project of moral improvement. This is where accounts that explain apparent weakness of will as a defect in intellectual character (as opposed to a more straightforward defect in the will) are insightful, much in the way that Aristotle's above-mentioned account does. If an agent is consistently motivated by the good and yet fails to act accordingly due to a lack of knowledge of what the good requires, this implies something more like frailty than impurity, since this agent would remain motivated by the good (not wayward desire) while incapable of following through on that motivation for reasons of intellectual character. There is a further illuminating analogy here with R.M. Hare's interpretation of weakness of will. The existence of cases like those mentioned by Hare renders Kant's picture of character increasingly plausible.

According to Hare, the failure of agents to do what they say they ought typically results from either impossibility to do so (whether physical or psychological), or a use of 'ought' in an insincere or off-color way.[36] The prospect of impossibility is helpful to the Kantian picture in at least two ways. First, Hare's discussion of divided personality provides a way in which an agent may fail to do what they ought to do while their will remains fundamentally good. The explanation Hare provides is that one part of the agent's personality issues a moral command to the other while the other is unable to obey this command because of a "recalcitrant lower nature" (Hare, 1963, p. 81). This explanation has the effect of retaining the strength of the command as well as the agent's ability to endorse it in their 'higher' self, even though their 'lower' self is incapable of obeying it. Such an agent may have a good will or character in Kant's primary sense of a fundamental motivation to do the good, and yet not have good character in the secondary, complete sense because this recalcitrant lower nature makes it impossible for them, at least for the time being.[37] If Hare's account is correct and such agents exist (and action in accordance with the good is impossible for them), it provides an explanation that further eases the puzzle surrounding Kant's dual conception of character. Though Kant believes this sort of change in character is possible, he holds that it takes time and requires knowledge of self to proceed, so in this sense it may not be possible for an agent to alter this recalcitrant nature all at once. This explanation also does not effectively handle the fact that mastery of oneself and the acquisition of character in the secondary sense is primarily a cognitive task and will involve epistemic states. This is better handled in a further point concerning the notion of impossibility.

This second point is that it may be that the agent of good character in the primary sense, whose fundamental intention is to do the good, is unaware of further patterns of thought that continue to linger on in their mind and self and direct them to not do the good, even after they have undergone the change of heart. Kant's example of allowing "apprentices in morality to

[36] If an agent intentionally does the latter, it is safe to say their will is not sufficiently good since an insincere and intentional utterance along these lines belies that they do not mean 'ought' in the proper way. Hare is probably right on this point though it is not of obvious relevance to Kant. For a summary of his position, see Hare, 1963, p. 82-83.
[37] It should be noted that Kant does not have room for the idea of a recalcitrant lower nature except insofar as it is not maxim governed and therefore not under an agent's control.

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

judge the impurity of certain maxims on the basis of incentives actually behind their actions" suggests something just like this (Kant, 6:48). It suggests that an agent must seek out and expunge beliefs, patterns of thought, and features of character that result from prior habituation in self-love, and that one who is inexperienced in morality cannot immediately detect these. Nevertheless, an agent can learn how to detect them and be successfully cured. Learning to detect such features of character and root them out is a matter of self-knowledge and a task on which the agent who has undergone the change of heart is embarked. It complements the Kantian virtue of sympathy, which is also cognitive in nature, since it involves the apprehension of facts concerning the well-being of others and one's duty toward them, and how to apply the moral law in particular circumstances. Knowledge of these features does not come immediately with the change of heart—it must be acquired through time—and adequate possession of this knowledge is necessary to make good character complete. Until this knowledge is acquired, the agent of good intention is ignorant of their existence and as a result, it is impossible for this person to eradicate them. In this way, it is possible that the Kantian agent may have a good will, or virtuous character in the primary sense, and not yet have virtuous character in the secondary, full sense.

This does not account for an agent who has knowledge of what the moral law requires and yet does not, as it were, activate that knowledge in a particular case. It would seem there are such cases in which desire for some course of action that is contrary to the moral law distorts an agent's belief that that course of action is in fact contrary to the moral law (for example, the belief that one's partner will be happier if they do not know the truth about a gambling addiction, so it is better to fabricate reasons for lost income). In some sense this agent can be said to be acting in ignorance of the good; however, it seems that they are culpable for this ignorance which makes it more difficult to believe that they have undergone a genuine change of heart. This appears to be a case of impurity, or conditioned autonomy as explained by Engstrom. This agent is not consistently acting out of moral commitment; their resolve is weak and their apparent change of heart episodic. Kant is attentive to this and cautions that we cannot know when the change of heart takes place.

> Assurance of this cannot of course be attained by the human being naturally, neither via immediate consciousness nor via the evidence of the life he has hitherto led, for the depths of his own heart (the subjective first ground of his maxims) are to him inscrutable. Yet he must be able to *hope* that, by the exertion of *his own* power, he will attain to the road that leads in that direction, as indicated to him by a fundamentally improved disposition (Kant, 6:51).

So, there is never a reason not to be vigilant concerning the moral status of one's maxims. The change of heart is something we hope in and strive to make evident by our effort toward moral action. Kant emphasizes in *The End of All Things* that this effort is unending as far as earthly life is concerned, saying that moral life even at its best "will always remain an ill compared with a better one", so the goal of moral satisfaction can only be posited beyond time and

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

understanding.[38] And he is alert to the dangers of self-deception, in the *Religion* cautioning that "one is never more easily deceived than in what promotes a good opinion of oneself" (Kant, 6:68). Perhaps he has this in mind when he identifies virtue with fortitude and the "resolve to withstand a strong but unjust opponent" (Kant, 6:380). In any case, this is quite different from an agent who, for example, begins with a disadvantage insofar as their mind and character are warped, bent away from the good, while their deeper desire and intention are to correct this. The change of heart entails an imperative to engage in this thoroughgoing work of correction: it signals the commencement of this project. And it takes time to get things right.

5. Conclusion

If my argument is correct, the virtues of sympathy and self-knowledge help to account for fully virtuous character on the Kantian picture. To acquire these virtues is to acquire a moral *Denkungsart* and therefore virtue in one's empirical character. This is one way to understand Kantian moral strength. If we look at this from the other direction, we can understand a lack of these virtues as a form of moral weakness. This provides the groundwork for a character-based account of Kantian moral weakness. According to this account, moral weakness is located in defective empirical character, namely the underdevelopment of sympathy and self-knowledge.

As mentioned at the outset, Hill has argued that weakness of will be understood as a character trait, rather than a feature of isolated acts only (Hill, 1998, p. 94-95). He suggests "we need to survey several aspects of the agent's history over time, including the degree of effort, the type of resolve, and the frequency and reasons for 'changes of mind'" (Hill, 1998, p. 107). He applies this suggestion to Kant, arguing that Kantian weakness of will is not found in a lack of power to do something, but in a vague resolution which weakens one's resolve to be moral while blurring the content of what morality requires. It "opens a door for self-deception, inattention, and special pleading that enable us to live with a genuine conflict of will, of which we are aware enough to be responsible but not enough to prompt us to change" (Hill, 2008, p. 223). These forms of weakness are less weaknesses of will than weaknesses of character—frailty—and among the deficiencies that sympathy and self-knowledge are meant to correct. For example, self-knowledge corrects self-deception while sympathy entails attention to the morally relevant details of others' lives. Since the acquisition of these virtues is one way to acquire moral strength, moral strength goes beyond the bare resolve to somehow will better. It is not merely a

[38] "Even assuming a person's moral-physical state here in life at its best – namely as a constant progression and approach to the highest good (marked out for him as a goal) – he still (even with a consciousness of the unalterability of his disposition) cannot combine it with the prospect of *satisfaction* in an eternally enduring alteration of his state (the moral as well as the physical). For the state in which he now is will always remain an ill compared to a better one which he always stands ready to enter; and the representation of an infinite progression toward the final end is nevertheless at the same time a prospect on an infinite series of ills which, even though they may be outweighed by a greater good, do not allow for the possibility of contentment; for he can think that only by supposing that the *final end* will at sometime be *attained*" (Kant, 8:335). And: "this is a concept in which the understanding is simultaneously exhausted and all thinking itself has an end" (Kant, 8:336).

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

matter of gritting one's teeth and doing the right thing. Of course, it may require the ability to do this when necessary, but it requires more besides this, including a clear, consistent, or firm understanding of what the moral law demands of one in the first place.

Conversely, moral weakness may be found in the absence of this understanding—the absence of these character traits. As I have argued, lack of sympathy and self-knowledge weakens one's ability to make the moral law efficacious in one's life and circumstances. Further, it requires resolve to cultivate these virtues in the first place, so an agent who fails to cultivate them may be deemed weak because of that failure. This is one way to understand Kant when he says the frailty of human nature is expressed when one incorporates the moral law into one's maxim, but that maxim turns out to be subjectively weaker when the time comes to act on it.[39] The subjective weakness of the maxim is reflected in the agent's failure (lack of resolve) to cultivate the virtues of self-knowledge and sympathy. This goes beyond the standard approach to moral weakness and strength which identifies that weakness (or strength) with instability (or stability) in one's basic commitment exclusively.[40] The puzzling passage on the change of heart in the *Religion* (Kant, 6:47-48) tells us that there is more to an agent's character than their basic commitment identified with their intelligible character. Instability in one's basic commitment is one form that weakness may take, while a lack of the above virtues is another. Again, it should not be surprising that human beings can be weak in more ways than one.

This contributes to a broader effort to identify moral strength and weakness with what might be called moral fitness, or a high degree of development in virtuous character. Merritt has argued for a similar account, pointing out that Kant's "idea that the virtuous and the holy have the same strength should give us pause about modelling the strength of virtue too closely on the strength of muscles" (Merritt, 2018, p. 187-188). She continues: "the strength that they share can only be the strength of practical reason—the strength of a cognitive capacity—however exactly this idea should be unpacked" (Merritt, 2018, p. 188). Merritt's account unpacks this in terms of cognitive skill. I have argued above for a different account of sympathy than the one provided by Merritt, but her account of strength is correct. It applies to the "whole package", including cognitive and motivational aspects alike (Merritt, 2018, p. 155). This parallels an earlier suggestion from Richard Henson that Kant has two accounts of moral worth, one for an ability to overcome contrary inclinations and another for general fitness in one's inclinations in the first place (Henson, 1979).[41] Similarly to how these accounts of moral worth are mutually consistent, the account of moral weakness I have argued for here is consistent with the standard account.

The difference is that the character-based account I have laid out here provides a more psychologically realistic means for moral improvement. Moving from moral weakness to

[39] "I incorporate the good (the law) into the maxim of my power of choice; but this good, which is an irresistible incentive objectively or ideally (*in thesi*), is subjectively (*in hypothesi*) the weaker (in comparison with inclination) whenever the maxim is to be followed" (Kant, 6:29).

[40] "We can say that someone's will is *stable* or *strong* if she tends not to alter her basic commitment too readily and she tends to revert back to it were it to change, while a person's will is *unstable* or *weak* if she tends to alter her basic commitment too readily and tends not to revert back to it were it to change" (Cureton, 2016, p. 71).

[41] A thorough discussion of Kant's definition of moral worth is beyond the scope of this paper.

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

strength is a matter of deepening one's sympathy and self-knowledge. There are clear ways that this can be achieved. For example, to cultivate sympathy one has "a duty not to avoid the places where the poor who lack the most basic necessities are to be found but rather to seek them out, and not to shun sickrooms or debtors' prisons and so forth in order to avoid sharing painful feelings one may not be able to resist" (Kant, 6:457). Spending time in the presence of those who are suffering can expand one's sympathies if one is attentive to the experience of those others and takes time to reflect on it. This is a matter of thinking into the place of the other to better understand how one may act for their well-being. As Kant says, to attain wisdom one must: "1) Think for oneself, 2) Think into the place of the other (in communication with human beings), 3) Always think consistently with oneself" (Kant, 7:200). Regarding self-knowledge, "sincerity in acknowledging to oneself one's inner moral worth or lack of worth are duties to oneself that follow directly from this […] command to cognize oneself" (Kant, 6:442). Fortunately, Kant does not believe we must go it alone and acknowledges the value of moral friendship, "the complete confidence of two persons in revealing their secret judgments and feelings to each other" (Kant, 6:471). Alongside self-examination and the moral education of children (as in the moral catechism, for example), the honest conversation and exhortation that is a part of moral friendship may contribute to the virtue of self-knowledge, both in oneself and in another.[42] The virtues of sympathy and self-knowledge provide a goal for these practices, focusing one's effort, which the standard account does not do as it defines moral strength as a matter of bare volitional resolve.

Finally, it is worth considering the objection that Kant's idea of freedom rules out the possibility of an account of moral weakness like this because in every moment an agent is free to decide between the moral law on one hand and self-love on the other, and there is therefore only one kind of weakness that manifests itself in a straightforward choice to not act on the moral law. This idea may seem to follow from a passage in the *Religion*, where Kant says:

> whatever his previous behavior may have been, whatever the natural causes influencing him, whether they are inside or outside them, his action is yet free and not determined through any of these causes; hence the action can and must always be judged as an *original* exercise of his power of choice (Kant, 6:41).

However, it is possible that while an agent may choose the moral law and commit to implementing it in their actions, they may fail to understand what it requires of them in the particular circumstances they find themselves in. They may likewise fail to observe in themselves a pattern of thinking or acting that is self-serving and causes action intended to be moral to derail when it meets the world. As I have argued, this stems from a failure to develop the two crucial Kantian virtues of sympathy and self-knowledge. These virtues are primarily cognitive, and their main purpose is epistemic. They provide an agent with information about themselves, others, and the world so that they may better fulfill their duty in real terms. It is therefore possible for an agent to choose the moral motive yet be unable to perform the proper moral action. The above account of Kantian virtue demonstrates how this problem is to be

[42] See Kant, 6:478-480 for the moral catechism.

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

overcome. It also avoids the possible challenge of the incorporation thesis, since the weak agent is not acting against their better judgment. Instead, the problem lies in their judgment itself, a feature of their character.[43]

Returning to moral strength, I have argued for an account of moral strength that is grounded in character. Rather than limiting the metaphor of strength to a matter of resolve or force in overcoming contrary desire, I suggest we may understand it as a high degree of development, integrity, or fitness in one's moral faculties. These faculties include most importantly sympathy and self-knowledge, cognitive virtues that Kant argues we have a duty to cultivate. Because the cultivation of these virtues requires resolve and continual effort, we may hold one who has them to be strong in the standard sense as well. Acquisition of this kind of moral strength provides an explanation of moral improvement in Kantian terms, as well an explanation of moral weakness. It also resolves a puzzling passage in the *Religion* (Kant, 6:47-48) concerning the difference between intelligible and empirical character and how an agent may possess good character in the former but not the latter sense. Beyond this, it is suggestive of a more general insight into Kant's idea of moral character, namely that it requires the cultivation of sympathy and self-knowledge, two cognitive and practical-epistemic virtues without which an agent cannot approximate virtue in the full sense.

[43] This is apparent in the introduction to the second part of the *Religion*, where Kant compares his position with that of the Stoics. "However, those valiant men [the Stoics] mistook their enemy, who is not to be sought in the natural inclinations, which merely lack discipline and openly display themselves unconcealed to everyone's consciousness, but is rather as it were an invisible enemy, one who hides behind reason and hence all the more dangerous. They send forth wisdom against folly, which lets itself be deceived by inclinations merely because of carelessness, instead of summoning it against the *malice* (of the human heart) which secretly undermines the disposition with soul-corrupting principles" (Kant, 6:57).

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

References

Allison, H. 1990. *Kant's Theory of Freedom.* Cambridge.

Aristotle. 1999. *Nicomachean Ethics*. Trans. T. Irwin. Indianapolis.

Bloom, P. 2016. *Against Empathy: The Case for Rational Compassion.* London.

Cassam, Q. 2019. *Vices of the Mind.* Oxford.

Charles, D. 1984. *Aristotle's Philosophy of Action.* London.

Cureton, A. 2016. "Kant on Cultivating a Good and Stable Will". In *Questions of Character*. Ed. I. Fileva. Oxford, 63-77.

Engstrom, S. 1988. "Conditioned Autonomy". In *Philosophy and Phenomenological Research* 48(3): 435-453.

Frierson, Patrick. *Kant's Empirical Psychology.* Oxford, 2015.

Hare, R. 1963. *Freedom and Reason.* Oxford.

Henson, R. 1979. "What Kant Might Have Said: Moral Worth and the Over-determination of Dutiful Action". In *Philosophy Review* 88(1): 39-54.

Hill, T. 1986. "Weakness of Will and Character." *Philosophical Topics* 14(2): 93-115.

Hill, T. "Kant on Weakness of Will". In *Weakness of Will from Plato to the Present*. Ed. T. Hoffmann. Washington, DC, 2008.

Holton, Richard. "Intention and Weakness of Will." *The Journal of Philosophy* 96, no. 5 (1999): 241-262.

Hume, D. 2007. *A Treatise of Human Nature Volume 1*. Eds. David Fate Norton and Mary J. Norton. Oxford: Oxford University Press.

Johnson, R. "Weakness Incorporated". In *History of Philosophy Quarterly* 15(3): 349-367.

Kant, I. 1996a. *Groundwork of the Metaphysics of Morals.* Trans M. Gregor. Cambridge.

Kant, I. 1996b. *The Doctrine of Virtue.* Trans. M. Gregor. Cambridge.

Kant, I. 1998a. *Religion Within the Boundaries of Mere Reason.* Trans. G. Giovanni. Cambridge.

Kant, I. 1998b. *The End of All Things.* Trans. A. Wood and G. Giovanni. Cambridge.

Kant, I. 2006. *Anthropology from A Pragmatic Point of View.* Trans. R. Louden. Cambridge.

Kant, I. 2012. *Anthropology Friedländer.* Trans. F. Munzel. Cambridge.

Merritt, M. 2018. *Kant on Reflection and Virtue.* Cambridge.

Morrison, I. 2005. "On Kantian Maxims: A Reconciliation of the Incorporation Thesis and Weakness of the Will". In *History of Philosophy Quarterly* 22(1): 73-89.

Palmquist, S. 2015. "What is Kantian *Gesinnung*? "On the Priority of Volition over Metaphysics and Psychology in *Religion Within the Bounds of Bare Reason*". In *Kantian Review* 20(2): 235-264.

Schapiro, T. 2011. "Foregrounding Desire: A Defense of Kant's Incorporation Thesis". In *Journal of Ethics* 15: 147-167.

Schapiro, T. 2009. "The Nature of Inclination". In *Ethics* 119: 229-256.

Seymour Fahmy, M. 2009. "Active Sympathetic Perception: Reconsidering Kant's Duty of Sympathy". In *Kantian Review* 14(1): 31-52.

DOI: https://www.doi.org/10.1007/s11406-022-00567-z

Sherman, N. 2014. “The Place of Emotions in Kantian Morality”. In *Kant on Emotion and Value.* Ed. A. Cohen. New York.

Vujošević, M. 2019. “Kant’s Account of Moral Weakness”. In *European Journal of Philosophy* 27: 40-54.

DOI: https://www.doi.org/10.1007/s11406-022-00567-z