

Kant's Metamaxims

Nichole Smith

October 8, 2019

1 Introduction

Kant has an idea of metamaxims¹ that is not immediately clear. David Sussman and Alison Hills give two competing accounts of his idea of a metamaxim. Kant is clear that every person chooses one metamaxim, the metamaxim is freely chosen, and one's choice of metamaxim is one's fundamental choice between being good or being evil. For Sussman, metamaxims are identical with all of a person's actions, just viewed differently. There are several ways of looking at your life. One way is to look at the sensible behavior. Another way is to look at the reasons that motivate you. And a third way is to look at whether you, on the whole, take self-love or morality to be a higher priority. On Sussman's account, the first and third are just two ways of looking at the same thing, and I argue that the second is as well.

Hills instead takes a dispositional account cashed out in terms of counterfactuals. For Kant, the choice of metamaxim is just a choice of priority of self-love over morality or morality over self-love. This choice is *the* choice between good and evil. One way to look at the question is to focus on what the prioritization is. Hills does this, and she argues that we should understand the prioritization as a disposition to take on maxims for the sake of self-love over morality or vice-versa.

The key dispute between these two accounts as considered here is whether an identity claim can be made between metamaxims, maxims, and actions. On the one hand, making

1. *Gesinnung*

the identity claim seems like it would come with some cost of explanatory power. On a dispositional account, we can say that the reason an action was chosen was for a maxim. And we can say that the reason the maxim was chosen was because of the metamaxim. There is something strange about explaining the maxims as chosen because of the metamaxim if the maxims are identical with the metamaxim. I will address this strangeness later. On the other hand, understanding metamaxims as identical with maxims and actions makes explaining some other things, like one's choice of a metamaxim, easier. If one's choice of metamaxim just is one's choice of maxims and choice of actions viewed with different perspectives, then while some things about each choice might be easier to harder to understand from the given perspective, we can also understand the whole as a rather down-to-earth and comprehensible idea.

Maxims play a key role in Kant's moral philosophy.² The categorical imperative explicitly assesses maxims. Actions are based on maxims. *Metamaxims* are mentioned in the *Groundwork*, but they become more important in the *Religion*.³ There are two particularly important roles that a metamaxim plays: One, a metamaxim somehow serves as the grounds for other maxims. Two, a person's choice of metamaxim is also her ultimate choice between good and evil. These two roles are related. Kant infers that we choose a metamaxim because we must understand ourselves as choosing between good and evil. The only available way to understand this choice is as a ground of all other maxims, because the concept of a metamaxim is a unification of all other maxims. Put in other words, one way we understand

2. Kant's account of morality is developed through the *Groundwork for a Metaphysics of Morals* (Immanuel Kant, *Groundwork of the Metaphysics of Morals*, 2nd ed., trans. Mary Gregor and Jens Timmerman (Cambridge: Cambridge University Press, 2012 (1785))), *Critique of Practical Reason* (Immanuel Kant, *Critique of Practical Reason*, 2nd ed., trans. Mary Gregor (Cambridge: Cambridge University Press, 2015 (1788))), *The Metaphysics of Morals* (Immanuel Kant, *The Metaphysics of Morals*, trans. Mary Gregor (Cambridge: Cambridge University Press, 1991 (1797))), and some other texts. It also relies on his account of transcendental freedom first explained in the *Critique of Pure Reason* (Immanuel Kant, *Critique of Pure Reason*, trans. Paul Guyer and Allen W. Wood (New York: Cambridge University, 1998 (1781/1787))). In his later text *Religion within the Boundaries of Mere Reason* (Immanuel Kant, *Religion Within the Limits of Reason Alone*, 2nd ed., trans. Theodore M. Greene and Hoyt H. Hudon (New York: Harper Torchbooks, 1960 (1793))), he further develops and partially revises his account, and in this text he really develops the idea of a metamaxim.

3. Alison Hills, "Gesinnung: responsibility, moral worth, and character," in *Kant's Religion within the Boundaries of Mere Reason*, ed. Gordon E. Michalson (2014), 79.

a person's myriad actions over time is as a unity. That unity is a maxim. For instance, someone who acts honestly all the time can be understood as having a maxim of honesty. In the same way, someone's choosing maxims that prioritize morality over self-love can be understood as a unity. That unity is a metamaxim.

In "*Gesinnung: Responsibility, Moral Worth, and Character*," Hills examines the idea of a metamaxim and how it works in Kant's system. That there are only two metamaxims and that one's choice of metamaxim is one's ultimate choice of good or evil demands further investigation. By the end of her article, she leaves some questions unanswered, and she says that Kant's account of morality is questionable because it takes a person's character to hinge on a choice of metamaxim. I will argue that some of these questions are not so mysterious and that Kant's account of morality is not so questionable if we adopt the understanding of metamaxims put forward by Christine Korsgaard and David Sussman.

In "Perversity of the Heart", Sussman explores and explains Kant's claim that "human beings are afflicted with an innate 'depravity' or 'perversity of the heart.'"⁴ He explains the first part of the *Religion* in which Kant explains the radical evil of all of humanity. In this explanation, he builds on Korsgaard's account of metamaxims as numerically identical with actions. From a merely intelligible point of view, we see only metamaxims. But from a sensory point of view, we see actions extended in time on a general trajectory of moral development or decay. It is this account that I argue in this paper can be extended to maxims in general and address the problems that Hills encounters.

As Hills does briefly respond to Sussman's article, I will first reconstruct Sussman's account of perversity, with primary focus on the two-standpoint account of maxims, including metamaxims. In his article, Sussman explains why Kant is led to have a concept of metamaxims in the first place. That reason just is to explain the perversity of the human heart. Given this motivation, we're left just to describe what a metamaxim and choice of metamaxim are. So, following this, I will retrace Hills's steps, focusing especially on the step at

4. David Sussman, "Perversity of the Heart," *The Philosophical Review* 114, no. 2 (2005): 153.

which she denies Sussman's account. She considers a few possible descriptions of metamaxims, but given Kant's motivation for using the concept of metamaxims, they do not make sense within Kant's framework. Finally I will defend Sussman's account and show how it answers to the problems Hills identifies.

2 Sussman's Perversity of the Heart

In Sussman's "Perversity of the Heart," he accomplishes at least two things. One, he provides a motivation for Kant's use of the concept of a metamaxim. If the idea of a metamaxim were, for example, just a demand of the principle of sufficient reason that choices of maxims have some grounds, then one could easily press the same argument on metamaxims to demand metametamaxims. A regress would easily follow. However, humanity has a problem, namely that we are evil. Metamaxims are key in understanding this evil. Two, given the choice of a metamaxim, he identifies and addresses a quasi-Calvinist worry. On many readings of Kant's idea of the (noumenal) choice of a metamaxim, while we freely choose our metamaxim, we have no sensible experience of this choice, and it seems wholly foreign. Calvinists believe in divine predestination. That is, God has already chosen everyone's ultimate destiny, so any attempt to be good is futile. The worry for Kant's theory is that if we have *already* chosen our metamaxim, then we have already chosen to be good or evil, so then nothing we do now in experience can make us good.

The project of this section is explaining why Kant says that we choose metamaxims and then explaining Sussman's description of a choice of metamaxim. Of course, Sussman does more than these two things, but our purposes only require these two. Sussman begins with Kant's bold claim that every human action is debilitated by radical evil. The rest of his article then investigates what this radical evil is supposed to be. Radical evil, for Kant, involves an active opposition to reason in our every action. Even the most virtuous human will find evil in even her most virtuous actions.⁵

5. Kant, *Religion Within the Limits of Reason Alone*, 6:32.

I say “human” rather than “person” because Kant makes a distinction between the two. A human being’s humanity is found in her being living and rational, but her personality is found in her being rational and responsible.⁶ This leaves open the possibility of non-human persons. This open possibility highlights the importance of our status as living, biological, and temporal persons. Evil distinguishes us humans “from [all] other possible rational beings.”⁷ Our being evil is a feature of our being living, biological, and temporal beings. Because we are not simply born with perfectly functioning rational faculties, our lives must be, at best, an ongoing struggle towards reason.

This process of human reason developing over the course of a life makes way for the understanding of maxims and metamaxims as, from one standpoint, atemporal imperatives, and, from the other standpoint, a series of sensible behaviors in time. Moreover, this development requires evil, so evil is “a necessary feature of any freedom that can be attained by a temporal creature that develops its reason out of a natural origin.”⁸

Kant’s moral philosophy has tight connections among morality, freedom, and reason. These seem to present a problem. If an act is only rational and free if it is morally permissible, then any immoral act must just be irrational and unfree. Then the act is merely something natural, beyond the control and responsibility of the actor. However, Sussman argues, Kant escapes this problem by only requiring implicit recognition of the moral law for an act to be minimally free. Kant subordinates the question of rational and free acts to the question of rational and free agents. If an agent recognizes the moral law, then she is free. So, she is responsible for her acts, and her acts are free just in virtue of her being free.⁹ This already sets the stage for metamaxims. An agent is ultimately good or evil. So, if the moral value of an action is determined by its maxim, the moral value of a maxim is determined by its metamaxim, and the metamaxim is¹⁰ the character of an agent, then the character of an

6. Kant, *Religion Within the Limits of Reason Alone*, 6:26.

7. *ibid.*, 6:21, but note in Sussman, “Perversity of the Heart,” 153 “all” is present.

8. *Ibid.*, 154.

9. *Ibid.*, 155.

10. The strict identity claim between metamaxims and character may or may not hold. For the purposes of this paper, they are close enough. The argument is for an agent-centric account of morality. If the

agent grounds the moral value of maxims and actions.

Kant considers three ways to explain human evil. The first is frailty, often called “weakness of will” today. The second is impurity, a sort of self-deception. The third is perversity, the radical evil that demands the concept of a metamaxim to understand. One might hope that frailty is enough to explain the apparent evil of humanity. Failing that, one might hope impurity (in addition to frailty) is enough to explain the apparent evil of humanity. Kant considers each of these and finds them to be insufficient. So, perversity is required. An account of perversity will then require an account of metamaxims. So, let us look at the arguments against frailty and impurity being sufficient to explain the evil of humanity.

Frailty is the first option to consider as an explanation of human evil. An instance of frailty is found whenever an agent recognizes the moral law yet chooses to act against it. But, while inclinations provide occasions for bad acts, they cannot determine our wrongdoing since we are free. Because practical reason is necessarily committed to the moral law, a frail-acting agent will not be able to make any rational sense of her bad action. It can at most appear as acting on a whim rather than a reasoned choice.¹¹ If one were to take the existentialist or nihilist line that there are no ultimate reasons, then everything might look like these sorts of whims. However, because Kant has a real moral law, there are some acts that can ultimately be done for reasons, namely moral acts. So, the category of frailty, or weakness of will, only explains some immoral acts.¹²

metamaxim is an essential determiner of the moral valence of a character, then the argument works.

11. Sussman, “Perversity of the Heart,” 155.

12. While someone who is normally principled and slips up can have her slipping up explained in terms of frailty, plenty of people actively embrace vice. As Sussman points out, episodes of frailty are usually followed by regret. Vices like envy, resentment, and bigotry tend to be embraced and endure throughout one’s life (ibid., 156). Again, seeing maxims and metamaxims as something expressed throughout one’s life will make more sense of this phenomenon. If you look at an episode of frailty in the moment, it can appear to be based on the self-love (inclination—Note that the use of “inclination” might also be suspect. Korsgaard points out (see Christine M. Korsgaard, “Morality as Freedom,” in *Kant’s Practical Philosophy Reconsidered: Papers presented at the Seventh Jerusalem Philosophical Encounter, December 1986*, ed. Yirmiyahu Yovel (1986), 29) that Kant’s use of “inclination” (*Triebfeder*) denotes the “candidate reasons among which she chooses.” Nonetheless, Sussman uses them interchangeably, and nothing in this paper hinges on the distinction.) and thus ultimately for the reason of that self-love. However, if the sort of thing an agent takes to be a reason does not ultimately include self-love but only the moral law then the episode is more clearly an inexplicable slip into a pure, irrational expression of agency.

In the *Religion*, Kant makes a useful move in his moral psychology away from that found in the *Groundwork* and second *Critique*: a distinction between the elective will (*Willkür*) and the purely legislative will (*Wille*).¹³ While there are varying accounts of the distinction,¹⁴ the key takeaway from both for Sussman’s argument is that the elective will need not always abide by the legislation of the legislative will. While the elective will ought to respect the authority of practical reason, it may instead “recognize inclination, in the form of what Kant calls ‘the maxim of self-love,’ as its highest concern.”¹⁵

At this point the driving problem enters the scene:

Enigmatically, Kant claims that we establish our fundamental disposition through a purely noumenal act that is somehow outside of time. He argues that each choice of disposition is necessarily unique and unchangeable, characterizing the agent from birth to death (and beyond).¹⁶ For each human being, there can be but one choice of disposition, a choice that is phenomenally expressed in a characteristic pattern of growth (or decay) in virtue throughout the whole course of her life.^{17,18}

13. Sussman, “Perversity of the Heart,” 156.

14. There are at least two conceptions of this division, most notably an elective conception and a rationalist conception. On an elective conception, the legislative will gives laws for motives of action. The elective will provides maxims. This is rather suggested by the *Religion* as your (freely chosen) character just is your choice between a good or evil metamaxim. Kant says in the *Metaphysics of Morals* that the will is neither free nor unfree, but *choice* is free and undetermined. Kant, *The Metaphysics of Morals*, 6:226 On this reading, the will is identified with the legislative will, and choice is identified with the elective will. Because the legislative will is just practical reason and so bound by the laws of reason, there is no election involved. In order for the will to operate, it needs to have a principle, so the will is in the business of choosing a principle so that principles can be made. See Korsgaard, “Morality as Freedom,” 29–48. On the competing rationalist conception of the two aspects of will, the legislative will is practically reasoning to a judgment on how to act. That is, the legislative will is the faculty that judges actions as good or not. The basic output of the legislative will is wishing, wishing being a judgment of an action or end as good, abstracted from one’s actual abilities. The elective will is then responsible for execution. It gives maxims, as maxims are conclusions about how to act. On both of these, we see that the legislative will, i.e. practical reason, does not settle what to do. However, the legislative and elective aspects of will are just two aspects of the same will. The legislative side must be merely intelligible, and the elective side can be found in empirical sensibility via the choices we actually make. Yet still, we might find empirical elements of the legislative will via moral feeling. When we feel bad for doing something bad, that can be explained by the work of the legislative will. We can also find merely intelligible elements of the elective will. In order to understand the choices, and maxims, we must take two points of view on them. On the one, sensible, point of view, they are expressed as actions in the sensible, temporal series of events. On the other, merely intelligible, they are understandable as maxims chosen atemporally. See Stephen Engstrom, “Reason, Desire, and the Will,” in *Kant’s Metaphysics of Morals: A Critical Guide*, ed. Lara Denis (2010).

15. Sussman, “Perversity of the Heart,” 156.

16. Kant, *Religion Within the Limits of Reason Alone*, 6:25.

17. *Ibid.*, 6:47.

18. Sussman, “Perversity of the Heart,” 157.

The problem is that every human being must choose between good or evil, and the single choice characterizes one's entire life. There is no possibility of change given that the choice is atemporal. This seems to force Kant into the position of all apparent moral progress or decay being merely epiphenomenal. The worries of Calvinism resurface.¹⁹ We are interested in moral theories in the first place because we want to be moral ourselves. We want to do the right thing or promote the good. A crucial question then is not just what is moral, but how to be moral. If a moral theory offers no conception of how improvement is possible, then it seems to be lacking in an important way. So, if the choice of metamaxim is the basic choice between being good and being evil, then the concept of the metamaxim better have room for an individual to improve over the course of her life. Otherwise Kant's moral theory will seem implausible.

As an individual's choice between the two possible metamaxims itself demands some explanation, Kant cites innate human propensities and predispositions. The propensities and predispositions are "necessary features of human nature that serve as the 'subjective determining ground' of one's fundamental disposition."²⁰ Kant distinguishes the two as follows: a propensity can be innate yet represented as not innate. We can think of propensities as acquired. Especially important in the case of evil, we can think of propensities as self-inflicted. Predispositions cannot be represented in this way. We can only think of a predisposition as something essential to being a kind of thing.²¹ Because Kant takes evil to be a propensity, he can say that being evil applies to a human being considered as a human being without saying that the evil can be inferred from one's being a human being.²² Kant finds, then, that philosophy is charged with the task of finding out what a rational creature must be like to be radically evil. That is, philosophy has the task of cognizing a priori what conditions we, despite our freedom and rational nature, must have to be evil and ergo to resist reason in our

19. Sussman, "Perversity of the Heart," 157.

20. Ibid.

21. Kant, *Religion Within the Limits of Reason Alone*, 6:29–30.

22. Ibid., 6:32.

every act.²³²⁴ Here we find a quite pressing motivation to understand Kant's conception of a *metamaxim*. The evil present throughout every human life needs to be explained in such a way that as we find ourselves in time we can still reasonably think that we have control over our future of being good or evil. It is Sussman's explanation of the *metamaxim* that I will argue succeeds in explaining how we can still reasonably think this way.

In addition to frailty not being sufficient to account for evil,²⁵ impurity also fails. Impurity, on Kant's account, is the habit of seeking motivation from self-love to support motivation from reason (morality) alone. For reasons Sussman explains that I will not rehash, it is ultimately an irrational form of self-deception that fails to explain evil entirely. What is important is that frailty and impurity are ultimately insufficient to explain human depravity.²⁶

To make sense of evil, then, we ought to briefly look back at Kant's distinction of humanity's animality, humanity, and personality. As animals, we are living beings. As human beings, we are living and rational beings. As persons, we are rational and responsible beings. All three of these are different predispositions.²⁷ That is, the features of living,²⁸ rationality, and responsibility must be conceived of as essential to our being human persons. As noted earlier, the distinction between personality and humanity reveals already the possibility of conceiving of rational beings abstracted from our status as living beings. From this standpoint, we can conceive of our responsibility to the moral law independently of our being in time with natural histories.

As animals we have some fundamental needs and impulses, such as drives to self-preservation, reproduction, and sociality.²⁹ It is here that we find our irrational starting point. From here Kant tells a story of development from irrational animals to rational persons.³⁰ By the end of

23. Kant, *Religion Within the Limits of Reason Alone*, 6:36,39n.

24. Sussman, "Perversity of the Heart," 159.

25. One might ask whether this accounting for evil is explaining or describing evil, but I take these two to be basically the same as far as this argument is concerned. Frailty is not sufficient to explain why we find humans to be evil, and frailty is in the same way not sufficient to describe the evil we find in humans.

26. Sussman, "Perversity of the Heart," 160–162.

27. Kant, *Religion Within the Limits of Reason Alone*, 6:26.

28. i.e. biological and temporal animal life

29. Sussman, "Perversity of the Heart," 162.

30. While animal inclination is not itself rational, it does have a rational structure. There are ends, such

the story of development, Kant says that “the idea of the moral law alone, together with the respect that is inseparable from it, cannot be properly called a *predisposition to personality*; it is personality itself”³¹ This is to say that we have a predisposition *to* personality, i.e. the idea of the moral law, but the idea of the moral law is not itself a predisposition. Kant indeed says that the “subjective ground” of “our incorporating this incentive into our maxims seems to be an addition to personality”³² and only on that ground is there a predisposition. There could then be something like a capacity to act from moral principle, but there is no predisposition to act from moral principle. Because of the absence of such a predisposition, we can act not according to moral principle.³³

This lays the groundwork for a history of human development. Obviously, humans do not come born with perfectly operating faculties of reason. In “Speculative Beginning of Human History,” Kant gives a possible account of the development from animal life, through developments such as honor and culture. At no point in the development can the project of attaining rational perfection ever be completed.³⁴ In order for the sort of evil that human beings are debilitated by to be possible, we must be the kinds of beings that “harbor a permanent antagonism to reason.”³⁵ As Sussman points out, in the first *Critique*, too, Kant sets up this problem. While we humans require education to the good, the educators are

as self-preservation, and means taken to those ends, like eating. This is to say that animal behavior is purposive. This lays the groundwork for rational behavior. Moving on to the human nature, Kant says that the predispositions to humanity fall under self-love, but they add an element of comparison (Kant, *Religion Within the Limits of Reason Alone*, 6:27). In the development of humanity from irrational to rational, we work our way up with this power of comparison. Rather than a human just seeing how good she has it, she compares how well she is doing with how well others are doing. This “comprises our essentially social and narcissistic drives” (Sussman, “Perversity of the Heart,” 163). Finally, our self-understanding as rational agents as such, i.e. in abstraction from our being living beings, shows us our predisposition to personality which “is realized by an implicit commitment to the moral law by which the agent is able to act not just purposively or rationally, but in a way that is fully free and autonomous” (ibid., 164).

31. Kant, *Religion Within the Limits of Reason Alone*, 6:28.

32. Ibid.

33. At the end of the section on the three predispositions, Kant describes some features of the three. The predisposition to animality does not have reason at its root; the predisposition to humanity has only instrumental reason at its root; and the predisposition to personality has practical reason itself at its root. All three are good both in that they do not resist the moral law as well as in that they demand compliance with the moral law. They are original, meaning that they are part of human nature in such a way that they cannot be escaped (ibid.).

34. Sussman, “Perversity of the Heart,” 165–171.

35. Ibid., 171.

themselves in the same basic position as the educated. Since no perfect rational being (i.e. God) is offering educational services, we rely on evil beings to deliver us from evil.³⁶ If we were able to completely escape evil, then we would be escaping what it is morally to be a human being. We can only make sense of our moral existence as in time and thus as in flux.

Sussman reiterates Korsgaard's suggestion that the (merely) "intelligible act that establishes our [metamaxim]³⁷ is not something numerically distinct from the ordinary empirical course of our practical lives."³⁸ This is to say that one's choice of good or evil, i.e. the prioritization of morality over inclination or vice-versa, is nothing other than the sum of one's actions over the course of one's life. The "noumenal choice" just is the sum of one's actions seen from a merely intelligible standpoint.

There are at least two immediately available gains from this reading. The first is that the Calvinist worry is now only present if it's present for the sum of one's life's choices anyway. While the metamaxim is indeed present from birth, its presence at birth is just our already having to conceive of a person as having a character. However, as something timeless, describing it as a choice that has *already* been made would be a mistake. If you take a walk, the walking is present as soon as you begin your walk, but this does not predestine you to arrive wherever you arrive, nor does it predestine you to walk the entire duration that you do. Rather, your acts of stepping in time determine where and how long you go. Calling it as a whole an act of or intention of walking is just to describe it from another standpoint.³⁹ Likewise, one is still in control of one's (im)moral behavior in time and can freely choose to be better or worse. But considered merely intelligibly, we conceive of the human as a fundamentally good or evil being.

The second gain is that this identity with the composition, so to speak, explains what a metamaxim is. Remember that the project of this paper just is to figure out what Kant's

36. Kant, *Critique of Pure Reason*, A 325.

37. Sussman uses the word "disposition" since "*Gesinnung*" can be translated as "disposition" or "metamaxim". He uses the two interchangeably. Hills's account takes a dispositional account of metamaxims to be distinct from Sussman's, so I follow her distinction for clarity.

38. Sussman, "Perversity of the Heart," 172.

39. Sussman uses a similar example. See *ibid.*, 172–173

conception of a metamaxim is. If whether one's being a good or evil human is dependent on her choice of metamaxim, then we surely want a clear idea of what a metamaxim is. The answer here just seems to be the sum, or composition, of one's actions. By "sum" or "composition" I do not mean to introduce a parts-whole ontology, but merely to note that there are two ways of looking at actions, one of which can be seen as subject to composition. Another example of this would be chairs. One way to conceive of my chair is four legs, a back, and a seat. Another way to conceive of that same chair is a comfy place to sit while typing. The parts listed in the first might not have a metaphysically-loaded composition relation to the end in the second. They could, but whether they do is irrelevant to this point. However, the material parts and the purposive object are numerically identical. Likewise, the actions needn't compose the metamaxim, but they do tell us what exactly the metamaxims are. One might now ask what the actions are, but the answer to that seems to be found by looking at humans move about in the world. As Sussman puts it, "the act of choosing a good [metamaxim] is to be found in a course of life that, as a whole, counts as making up one's mind to live morally and autonomously."⁴⁰ The making up of one's mind is found in one's continual progression from the irrational and immoral towards the rational and moral. Thus it is by the light of figuring out evil that we can most easily understand the metamaxim. ⁴¹

To conclude this section, on Kant's account, every human will is necessarily debilitated by radical evil. That is, everyone harbors an innate antagonism to reason and morality. This is simply the necessary way of existence for beings that are both rational and living. As finite organisms that must develop through time, we must have some sort of innate evil to allow for the possibility of development. Understood as a temporal unity, we must conceive

40. Sussman, "Perversity of the Heart," 172.

41. For further clarity, we can consider Sussman's account of a wicked person. Because of the timelessness of the metamaxim to good or evil, it must be "implicit in every point of my development, or it never really comes into being at all" (ibid., 173). There is an archetypal biography of the good person that outlines a progression toward true moral autonomy. The bad person at some point radically diverges. Because of the evil inherent to every human, all humans fail to live up to the idea of the most perfect possible human. However, the bad person will at some point stop at a certain point of development. Via means such as self-deception, the really wicked person will stop progressing towards moral autonomy and instead find ways to move into just another form of the same moral depravity (ibid., 174).

of a life as either ultimately prioritizing the moral over self-love (thereby taking on a good maxim) or else ultimately prioritizing self-love over the moral (thereby taking on an evil maxim).⁴² But, understanding a life as a temporal unity is just one way to understand it, and we can just as well understand it as a series of sensible actions. That is, there are multiple ways to understand the moral status of a human life. As a maxim, the unity is very clear, so the sharp distinction between good and evil is easy to understand. As a series of actions, becoming better or worse over time, the freedom is easier to understand.

With this conception of Kant's conception of the maxim in place, I will now turn to Hills's argument against it. In this section, we have found that Kant needs the concept of a maxim to explain human evil. Frailty and impurity do not provide an adequate account of human evil, and perversity is needed. In turn, perversity can only be understood as an evil choice of maxim. Following this, the question is what the choice of maxim is. The initial worry is that it is something mysterious that leads to a quasi-Calvinist worry of predestination. Sussman's account of Kant's account of maxims as identical to actions helps with this worry. So long as we can already understand our sensible actions as free, if they are identical with our maxims, then we can understand our maxims as free.

3 Hills's Maxim

In "*Gesinnung: Responsibility, Moral Worth, and Character*," Hills is, as I am, primarily concerned with understanding what Kant's idea of a maxim (*Gesinnung*) is. Rather than approaching the question from radical evil, she instead approaches it, perhaps more clearly, from maxims. Once again, the motivation for understanding Kant's idea of a maxim is to understand his idea of radical evil and his claim that humans have a natural propensity to choose evil.⁴³ By this light, Hills is after basically the same things as Sussman. However, her approach from maxims to maxims leads to some confusions. Since Kant

42. Sussman, "Perversity of the Heart," 174–175.

43. Hills, "Gesinnung: responsibility, moral worth, and character," 79.

is using metamaxims to explain radical evil, we should expect metamaxims to be fitting for that task. While they may fulfill other roles, this is something of an added bonus.

Hills comes to directly oppose Sussman's understanding of metamaxims. However, she ends up with more questions, and a less palatable verdict on Kant's account of moral character. In this section I will reconstruct her argument about what Kant takes a metamaxim to be, especially the parts that most directly engage with Sussman's. After this, we will be in a position to contrast the two and see how Sussman's account both explains more and makes Kant's account more plausible.

To approach metamaxims from maxims first requires an understanding of maxims. Kant explicitly says that a maxim is "the subjective principle of willing,"⁴⁴ contrasting with the objective principle, law. In the *Groundwork*, Kant works through several examples of maxims and how to test them with the categorical imperative.⁴⁵ In each example we find that they are basically the reasons why people do things. For instance, take someone who is in need of money. Acting on a maxim of honesty, she will only take a loan if she believes that she can repay the loan. Acting on a self-interested maxim, she might take the loan regardless. Her own reason is important here. As the subjective principle of willing, we are not looking for some causal explanation of why she does or does not take a loan, but rather what reasons she acts on.

Curiously, each maxim takes a future-tensed form. E.g. "when I believe myself to be in need of money, I shall borrow money, and promise to repay it, even though I know that it will never happen."⁴⁶ Hills also points out that they must be expressible as propositions, be described as actions, and accompany actions with reasons for action.⁴⁷ This description rather fits natural laws as well. They are usually put as future-tensed propositions that describe sensible phenomena (like actions) with reasons given. Natural laws are put in terms of some antecedent conditions leading to an event. Maxims are put similarly, as, ideally,

44. Kant, *Groundwork of the Metaphysics of Morals*, 4:401.

45. *Ibid.*, 4:422-423.

46. *Ibid.*, 4:422.

47. Hills, "Gesinnung: responsibility, moral worth, and character," 81.

whenever some conditions obtain, some action (event) obtains. Of course, since humans are not entirely rational beings, maxims do not always work out perfectly in the world, but they do take this form. This helps make some sense of autonomy being the will's "property of being a law to itself."⁴⁸ We can understand the laws of nature easily enough as the way natural phenomena must follow other natural phenomena. Likewise, then, we can understand the phenomena of the will following other phenomena of the will. This is further confirmed in the *Religion*, as Kant equates incorporation of incentive into one's maxim with having "made it into a universal rule for himself, according to which he wills to conduct himself."⁴⁹ Presumably this process of incorporation needn't be conscious. Hills notes that according to the incorporation thesis, we only act on an incentive (self-love) when we "freely choose to incorporate it into a maxim."⁵⁰ Thus our free choices just are those choices to incorporate certain incentives into the law-like maxims for our actions. I argue that these maxims are our actions, but Hills does not agree with this. When I say that maxims are our actions, I am saying that "maxims" and "actions" are extensionally referring to the same collections of things. However, we consider them from different perspectives, and on these two perspectives they appear as maxims and actions. As maxims, they are basically rules that we follow. As actions, they just are the behaviors we observe.⁵¹

According to Hills, there are three roles for a metamxim to play in Kant's account: the ground of moral responsibility, reason for action, and character. I will investigate each of these in further depth in this section. Kant explicitly describes the metamxim as a common ground of all maxims and itself a maxim.⁵² So, whatever account of maxims Kant has better apply to metamaxims as well. A rather confusing detail is that one's choice of metamxim can change. That is, one can choose an evil maxim and then choose a good

48. Kant, *Groundwork of the Metaphysics of Morals*, 4:447.

49. Kant, *Religion Within the Limits of Reason Alone*, 6:24.

50. Hills, "Gesinnung: responsibility, moral worth, and character," 81.

51. One might note here that in saying this, I am committing to someone only having the maxims that she actually acts on. For instance, someone only has a maxim of honest if she actually is honest. Someone who claims to be honest but actually deceives very frequently does not actually have a maxim of honesty.

52. Kant, *Religion Within the Limits of Reason Alone*, 6:20.

maxim.⁵³ According to Kant, “How it is possible that a naturally evil human being should make himself into a good human being surpasses every concept of ours.”⁵⁴

Moral responsibility is the first role for metamaxims that Hills considers. Whether a human’s doing a good or evil action only reflects on her character inasmuch as the action can be traced back to a good or evil metamaxim because the metamaxim is the ultimate ground of moral responsibility.⁵⁶ Hills notes that on Kant’s account, you’re only morally responsible for your action if the action is chosen on the basis of a freely chosen maxim. That maxim, however, could be chosen on the basis of an underlying metamaxim. There are two ways this could go. The first takes the choice of a maxim as an action, and the second considers the choice of maxim as based on reasons. If choosing a maxim is a free act, then that choice must be based on another freely chosen maxim. If choosing a maxim is better considered as it is based on reasons, then for the choice of a maxim to be morally evaluable, it better be based on reasons, and those reasons can only be supplied by a further maxim. So, whichever way we look at it, moral responsibility demands a hierarchy of maxims somehow grounded in maxims.⁵⁷

Kant’s demand for free choices to be undetermined, at least by elements of the causal sequence of nature, to qualify as morally relevant commits him to choices that are not causally determined. This shift to noumenal choices, according to Hills, becomes “even more mysterious one we have a hierarchical model.” The particular problem she brings out is that the grounding of a maxim in a metamaxim cannot be a determining grounding, as that would eliminate the possibility of freedom.⁵⁸

53. Hills, “Gesinnung: responsibility, moral worth, and character,” 82.

54. Kant, *Religion Within the Limits of Reason Alone*, 6:44-45.

55. While Hills says that Kant says it is not something that he can explain (Hills, “Gesinnung: responsibility, moral worth, and character,” 82), Kant infers from the possibility that there must be “a germ of goodness left in its entire purity” (Kant, *Religion Within the Limits of Reason Alone*, 6:45), and then describes the *process* of restoration of the original predisposition to good. A human being is only good in “incessant laboring and becoming[. . .] the change [from evil to good] is to be regarded only as an ever-continuing striving for the better” (ibid., 6:48). This is a problem I will return to in the next section.

56. Hills, “Gesinnung: responsibility, moral worth, and character,” 82-83.

57. Ibid., 84.

58. Ibid., 86.

Another apparent problem is the infinite regress. As noted, the metamaxim is itself a maxim. So, like the choice of other maxims, for the choice of metamaxim to be free, it must be on the basis of some freely chosen maxim, a “meta-meta-maxim” as Hills calls it. Then the meta-meta-maxim demands a freely chosen meta-meta-meta-maxim, and so on. She initially proposes that this regress can be avoided by denying that the choice of maxim is an action. But if that’s the case, then we’re left without a need for the metamaxim. If we step away from the consideration of a maxim as a ground of action and instead look at it as a reason for action, we still get a similar regress. While Kant recognizes the regress problem, he does not address it. Hills claims that Kant does distinguish the choice (“deed”) of metamaxim from the choice of other maxims,⁵⁹ and I take it she is referring to Kant saying, “by the concept of a propensity is understood a subjective determining ground of the power of choice *that precedes every deed*, and hence is itself not yet a *deed*.”⁶⁰ He further clarifies that there are two senses of “deed” (or, action) that might be applied, and they correspond to original sin and derivative sin.⁶¹

On the reason-seeking side, Hills looks at when questioning grounds makes sense. Indeed, in most cases, when someone cites a reason for action, you can ask, “why do you take that to be a reason for action?”⁶² The answer to that could be questioned the same way, but according to Kant, the line of questioning must bottom out in morality or self-love. The metamaxim, Hills argues, must be different from other maxims, then. This is precisely because the choice of metamaxim is a choice of what counts as reasons. She notes, though, that this leaves the question of how to change a metamaxim mysterious. If, for instance,

59. Hills, “Gesinnung: responsibility, moral worth, and character,” 87.

60. Kant, *Religion Within the Limits of Reason Alone*, 6:31, emphasis original.

61. Original sin is the original choice of humanity to do evil. Derivative sin is all of the day to day sins we commit as a result of humanity’s fall into sinfulness. Likewise, the choice of metamaxim is the sort of choice to be the kind of person that does evil or good. And actions are the day to day evil or good acts we commit. In the sense of original sin, the propensity to evil that is the evil metamaxim (as both are the ground of maxims) is indeed not an action. However, in the sense of derivative sin, it is an action. The sense of derivative sin is sin found in time. That is, the sort of temporal choice that lets us see evil in the phenomenal world. Original sin cannot be found anywhere in time, and so it is not an action in that sense. However, we are still entirely responsible for it, so in that sense it is an action. See *ibid.*, 6:31–32.

62. Hills, “Gesinnung: responsibility, moral worth, and character,” 88.

you take self-love to take priority over morality, then there is no apparent reason to invert this.⁶³ The question of how the inversion works, then, must be left open.

Reasons for action is the second role for metamaxims that Hills considers. The role is similar to that of moral responsibility. Your actions reflect on your moral character only inasmuch as your reasons were morally good. Your metamaxim is the ultimate reason for all of your actions, so it ultimately determines the moral worth of any action.⁶⁴⁶⁵

A conception of character is the third role for metamaxims that Hills considers. The metamaxim is the determiner for whether a person is good or evil.⁶⁶ By looking at character, we can judge agents rather than actions, and the metamaxim is the vehicle for this.⁶⁷ The first question Hills poses is whether we can “explain the moral evaluation of a person entirely in terms of good or bad reasons for action?”⁶⁸ Perhaps someone is evil just if all of her actions for which she is morally responsible are done for evil reasons. Or perhaps the bar is set at almost all actions being for evil reasons. Or perhaps the bar is merely at a majority being for evil reasons. While any of these might be refined, they all have the problem, she argues, that it becomes impossible to explain someone’s doing evil with her being evil, or doing good with her being good. Presumably, we would like to be able to explain someone’s actions with her character. The metamaxim as character serves this role.⁶⁹

Immediately this identification of metamaxim with character raises further questions. One standard way of understanding character is as a disposition to perform certain kinds of actions. She asks whether we can understand one’s metamaxim as a disposition. In

63. Hills, “Gesinnung: responsibility, moral worth, and character,” 88–89.

64. *Ibid.*, 83.

65. While the first role of the metamaxim establishes moral responsibility, the second role establishes moral worthiness. The same basic problems come up but in the examination of good action rather than evil action (*ibid.*, 89–91). In these first two roles we look to an agent’s metamaxim to explain why she does good or evil things and also why her doing those things is good or evil. One might also take a metamaxim to condition what an agent takes to be a reason for a maxim. This is compatible with the account of metamaxims thus given. The important point here is that we look to metamaxims to answer questions about maxims, such as why they are chosen and whether they are good maxims.

66. This might be better stated as the choice of metamaxim just is whether a person is good or evil.

67. Hills, “Gesinnung: responsibility, moral worth, and character,” 83.

68. *Ibid.*, 91.

69. *Ibid.*

that case, the other maxims are “manifestations” of that disposition. If that’s the case, we must ask whether one can choose a (first-order) maxim contrary to the metamaxim. Since there are only two possible dispositions (metamaxims), this is just to ask whether sometimes a good person can act badly, or a bad person act well. That seems very possible. But, if the relationship between a disposition and its manifestations is causal, then the relationship between metamaxim and first-order maxim must be causal since the metamaxim is a disposition manifested in first-order maxims. But the metamaxim is noumenal so it cannot be causal. So, so long as we are considering the metamaxim from the noumenal standpoint, it must not be a causal sort of disposition. Moreover, as established earlier, the maxims are all to be freely chosen. The metamaxim then must not determine them.⁷⁰

In resolving this question of how the metamaxim could be character, Hills next addresses Sussman’s explanation of having a metamaxim being “more like living a certain kind of life.”⁷¹ She says that while the account is appealing and elucidating, it does not fit with the features Kant ascribes to the metamaxim. Here I will address the two she puts forward.⁷²

The first: “that it is a ground of individual maxims seems to require that the *Gesinnung* is separate from those maxims, but it is not clear that there is sufficient separation here.”⁷³ But consider forms of grounding that we are familiar with that do not require much separation. Consider a dance.⁷⁴ It consists of many moves. Each of those moves is what it is in part because of the dance it is a part of. That is, the dance moves are grounded in the dance as a whole. This grounding relation does not require separation—in fact it requires the identity of the sum of the dance moves with the dance. If one is willing to accept that maxims can be identical with actions in this way and that maxims are the ground of actions, then the identity of maxims and metamaxims (their ground) is no problem. Moreover, the grounding relation has to do with a metamaxim that is only understood from the noumenal standpoint as a

70. Hills, “Gesinnung: responsibility, moral worth, and character,” 92.

71. Ibid.

72. Ibid., 93.

73. Ibid.

74. This example is discussed at length in Kenneth L. Pearce, “Foundational Grounding and the Argument from Contingency,” *Oxford Studies in Philosophy of Religion* 8 (2017).

metamaxim, as affirmed in the rejection of the maxim-as-causal-disposition account of the identity of maxim with character. Thus the grounding relation between maxim and maxim must not be one of efficient causation. So, this problem turns out to hold no weight against Sussman’s account.

The second: “according to Kant, in order for action based on the *Gesinnung* to be imputable to you, you must have chosen your *Gesinnung*.”⁷⁵ It is not clear that there is or could be a genuine choice of a life as a whole.”⁷⁶ However, as evident in the previous section of this paper, I take the genuine choice of a life as a whole to be one of the major payoffs of Sussman’s account. So, here I will clarify. One way to see the choice of a life as a whole is as the choice of each component. If every action is chosen freely and the sum of all actions is the whole, then the whole is chosen freely. What more one could demand than $\forall xF(x)$ —where $F(x)$ denotes x ’s being chosen freely, and x is quantified over the domain of actions in one’s life—is entirely mysterious. Perhaps what is imagined is one being able to stand outside of one’s own life, browse the options of possible lives, and select all actions at once. But this picture depends on the choice being something temporally prior to every choice. That is, some additional choice the comes before and determines all (other) choices. But if the choice of maxim is timeless, then it cannot be something before every other choice, as that would place it in time.⁷⁷

I will return to further defend Sussman’s account, but first I will finish reconstructing Hills’s argument so that her positive account can come out. Her next move is to suggest that the right account takes “seriously that the maxim is a kind of maxim.”⁷⁸ How this diverges from Sussman’s account is unclear given both maxims and maxims come

75. Kant, *Religion Within the Limits of Reason Alone*, 6:25,32.

76. Hills, “Gesinnung: responsibility, moral worth, and character,” 93.

77. Perhaps there is a risk of triviality here. After all, on Sussman’s account, the overall trajectory is what we see in the maxim. However, a trajectory is just a series of instances. We can consider the analogy to natural law again. While true that objects fall to the earth at a certain acceleration, the data we have is a series of moments with objects at a series of locations. Each location at each time is determined by the previous location and time. Likewise, then, we have a series of actions at a series of times. Their overall trajectory can be described, but each moment can be described as well, and each in terms of each other.

78. Hills, “Gesinnung: responsibility, moral worth, and character,” 93.

out to be identical with some part of a life, maxims merely smaller parts.⁷⁹ Moreover, though, the kind of maxim is the kind with a content that is to take certain considerations as reasons for actions. This reason-selecting maxim can then lead to choices of first-order maxims. She illustrates this by considering the metamaxim-maxim relationship in terms of counterfactuals. Her example: Say you have chosen to put self-love over morality. Then you probably have a maxim to take out loans on false pretenses. But, if you had chosen to put morality first, then you would not have that maxim. This counterfactual explanation has the virtue of giving explanatory power to the relationship between metamaxim and maxim while the first-order maxim is also not determined by the metamaxim.⁸⁰

An objection that she considers is that Kant admits that people who put self-love first and people who put morality first often end up performing the same actions. So, counterfactual dependence does not seem to be able to do the explanatory work. However, only the same actions are done, and just because the actions are the same does not mean that the (first-order) maxims are the same.⁸¹ In the examples Kant provides in the *Groundwork*, people doing the right thing for the wrong reasons turn out to be condemnable for their reasons because the right action is only contingently the result of the bad maxim. Kant demands a good maxim necessarily lead to good actions. The question, then, is whether a bad metamaxim can lead to a good maxim. If not, then indeed a difference in metamaxims can explain a difference in maxims. Hills notes that on occasion a bad person can choose a good maxim. So the explanatory power is unclear. If there is similarly this contingent possibility of something bad reasons nonetheless leading one in the right direction, then so long as a good metamaxim necessarily produces good maxims, outside of instances of deception and frailty, the explanatory connection is at least as strong as between first-order maxims and actions.⁸² Her move is to put the explanatory connection between metamaxims and maxims

79. “Smaller” here need not mean “shorter in duration”. One’s first-order maxims may all last an entire lifetime but only apply in certain domains. E.g. “Don’t lie” and “Don’t steal” apply to one’s entire life, but generally two different aspects of it.

80. Hills, “Gesinnung: responsibility, moral worth, and character,” 93.

81. *Ibid.*, 94.

82. *Ibid.*, 94–96.

on the same level as between maxims and actions. By doing so, she makes any problems one might have with her account on these explanatory grounds have to apply to both the first-order maxims and the metamaxims.

This leads finally to Hills's objection to Kant's account of morality. Because one's moral character comes down to, and is in fact identical with, one's metamaxim, feelings and emotions are left out of any possible central role. Because they are not freely chosen, we cannot be praiseworthy or blameworthy for them.⁸³ She thinks that it is "questionable whether the moral evaluation of a person can be based entirely on the choice of a metamaxim, rather than partly on one's non-cognitive attitudes, emotions, and feelings."⁸⁴

4 The Advantages of Sussman's Account

While Hills's account of metamaxims explains some things, it leaves some important questions open. Also, as Hills says, Kant's moral theory is implausible on her account. So, if an account can both address the problems while retaining the strengths of Hills's account, then we should favor that account. In this section, I argue Sussman's account does this. One problem in Hills's account is that the process by which one can become good remains mysterious. While she appeals to Kant's admission of mystery, Kant makes statements in the *Religion* supporting some positive claims about the possibility of becoming better. A second potential problem comes from that grounding cannot be determining. If Sussman's account is to accurately portray Kant, one's choice of metamaxim must not make one's choice of maxim unfree. However, I will argue that Sussman's account better this does than Hills's. A third potential problem comes from the infinite regresses. While she appears to cut off the regress by considering the metamaxims as special in not needing further grounding, some explanation for this is desirable. I will explain how a modest expansion of Sussman's account can explain the specialness of the metamaxims. The final problem I will address in this

83. See, e.g., Kant, *The Metaphysics of Morals*, 6:408

84. Hills, "Gesinnung: responsibility, moral worth, and character," 96.

section is that Kant's account of metamaxims and moral character do not adequately take into account emotion and non-cognitive attitudes. As Hills does not explain this objection in much detail, I can only explore it a little, but I will aim to provide a sort of emotional justification for Kant's metamaxim-based judgment of moral character. Finally, since the strengths of Hills's account derive from the counterfactual-dispositional understanding of the metamaxim-maxim relationship, I will show that the counterfactual understanding is preserved under identity.

Concerning the first problem, that the process by which one can become good remains mysterious: Despite Hills's claim that the possibility of becoming good after being evil is mysterious even by Kant's admission, Sussman's account both makes sense of it and also has textual support in the *Religion*. While Kant does say that how a "naturally evil" human being could make herself good is beyond any concept we have,⁸⁵ this only requires that the method one uses to move from evil to good is mysterious. Just because we cannot explain how something happens does not mean that it cannot happen. Kant does, however, have some speculative ideas of how humans might develop toward the good. Where can the mystery be, then? If it's not in the possibility of a path from evil to good, it must be in the motivation to become good. That is, we cannot conceive of why anyone entirely wicked would ever decide to embark on the journey to goodness. Necessarily an evil person must decide to remain evil.

If Kant just left the description of humans as entirely evil, then we would indeed be left in mystery. He does not. Instead, he infers from the possibility of movement from evil to good that in evil humans there must in fact be some goodness "in its entire purity."⁸⁶ Given this goodness, he has a starting point for the *process* of restoration to the original predisposition to good.

As cited above, Kant says that a human being is only good in "incessant laboring and becoming[. . . and] the change [from evil to good] is to be regarded only as an ever-continuing

85. Kant, *Religion Within the Limits of Reason Alone*, 6:44-45.

86. *Ibid.*, 6:45.

striving for the better.”⁸⁷ This incessant laboring and becoming is quite apparent on Sussman’s account. Someone with a good metamaxim just is someone who labors throughout her life to grow the small purity of goodness in her as much as possible. Inversely, someone with a bad metamaxim just is someone who does not work to grow the goodness and instead lets the evil propensity reign supreme.

Concerning the second problem, arising from that one’s choice of metamaxim must not make one’s choice of maxim unfree on Kant’s account: Hills claims that the counterfactual dependence is not determining. While this is possible, it does not have much explanatory power. In some cases, a different metamaxim would lead one to have different maxim, though which cases is left open. If the cases were not left open, then there would be a determining relation.

On the other hand, if the metamaxim grounds maxims inasmuch as the maxims are just parts of the metamaxim, then the metamaxim has explanatory power inasmuch as one can tell which maxims there are on account of the metamaxim, but the metamaxim does not determine the maxims any more than the maxims, and actions for that matter, determine the metamaxim. Because the three are identical, each is just as free as the others. Again the account that takes metamaxims, maxims, and actions to ultimately be the same things makes more sense. Because we can understand actions as free, we can understand anything that is identical with actions as free.

Concerning the third potential problem, from infinite regresses: Hills cuts off the regress by considering the metamaxims to not be actions of the same kind and therefore not dependent on further maxims. Kant does categorize the metamaxim as akin to original but not derivative sin, and therefore an action qua responsibility but not qua empirical choice.⁸⁸ But some explanation of why the metamaxim is unique is still desirable. If actions and maxims and metamaxims are numerically identical things considered from different aspects, we should ask why we don’t find another standpoint from which we see metametamaxims.

87. Kant, *Religion Within the Limits of Reason Alone*, 6:48.

88. *Ibid.*, 6:31–32.

To answer this, we need to consider the line of thought leading to each of the three we do find. Actions are found by sense; we look with our sense organs and see actions. We may ask why one does a certain action, and in order to understand this, we must conceive of maxims. Finally, we might ask whether a person is fundamentally good or evil, and, as argued in the first part of the *Religion*, we must conceive of metamaxims to answer that question. At this point we are simply out of questions. While there is no reason to think the idea of a metametamaxim is impossible, we do not simply postulate further orders of maxims without a question to answer.

We might still try to press the regress directly, asking what grounds the metamaxim, or what explains the metamaxim, so considering another angle might help. Maxims are, on Sussman's account, a sort of unification of actions. Someone who acts for the sake of honesty, for instance, can have her honest acts considered as a unified entity in the form of a maxim to be honest. Likewise, someone whose maxims are, on the whole, prioritizing morality over self-love, is understood to have a good metamaxim. The metamaxim is a unification of maxims. So, a maxim explaining actions is also a unified understanding of those actions. A metamaxim explaining maxims is a unified understanding of those maxims. However, because an individual only chooses one metamaxim, there is nothing left to unify. The end of explanation is not arbitrary but rather because there is in fact nothing left to explain.

Concerning the final problem, that Kant's account of metamaxims and moral character do not adequately take into account emotion and non-cognitive attitudes: Hills does not go into detail on this objection, but the main thrust of it is that accounts of character like Aristotle's take into account states such as the emotional. However, because we do not freely choose these things, we cannot be morally responsible for them.⁸⁹ Now, if this problem is just fundamental, and one wishes to deny any moral theory that does not make one responsible for that which one does not have control over, then of course Kant's theory is out from the

89. Kant, *The Metaphysics of Morals*, 6:408–409.

start. However, part of the problem here might just be the mysteriousness of the *metamaxim*. If it is something understood only as a thing we can only conceive of as a sort of truth of our goodness or evil which our first-order reasons for action counterfactually depend on, then it does seem rather divorced from our lives. But, Sussman's account offers an explanation of *metamaxims* and therefore our foundational character in a non-mysterious way.⁹⁰

If by "the most prominent conception of moral character" that is in conflict with Kant's moral theory⁹¹ Hills means an account of moral responsibility which holds us responsible for things we have no control over, then I do not know how to respond as I do not know how we can be morally responsible without some sort of causal or otherwise free responsibility. If she means an account in which our responses to our emotions and any free actions that might influence our future emotions, then the *metamaxim* captures this on Sussman's account as it captures all actions. If she instead means some account of how to live happiest, then the problem is not between the *metamaxim* account of moral character and the opposing moral theory but rather the conception of what a moral theory is supposed to capture. The problem is not that Kant is wrong about any role the *metamaxim* plays but rather that Kant's idea of a moral theory only includes our freely-made choices. He could have a theory that captures how to live a happy life; it just falls outside of what a moral theory is.

Finally, we should consider the ability of Sussman's account to take on the strengths of Hills's. That is, the ability of the identity of *metamaxims* with *maxims* to capture counterfactuals. Just as a strict matter of logic, if a collection just is the composition of a collection of collections, then a change in the first collection necessitates a change in the latter collections, and vice-versa, if the identity is upheld. Then all identity relations capture counterfactuals, include that of the *metamaxims* to *maxims*. Moreover, as explained earlier in this section, while Sussman has to leave open how the counterfactual relation obtains without being determining, the identity relation can ground a strict counterfactual while not being determining.

90. Hills, "Gesinnung: responsibility, moral worth, and character," 93.

91. *Ibid.*, 96.

References

- Engstrom, Stephen. "Reason, Desire, and the Will." In *Kant's Metaphysics of Morals: A Critical Guide*, edited by Lara Denis. 2010.
- Hills, Alison. "Gesinnung: responsibility, moral worth, and character." In *Kant's Religion within the Boundaries of Mere Reason*, edited by Gordon E. Michalson. 2014.
- Kant, Immanuel. *Critique of Practical Reason*. 2nd ed. Translated by Mary Gregor. Cambridge: Cambridge University Press, 2015 (1788).
- . *Critique of Pure Reason*. Translated by Paul Guyer and Allen W. Wood. New York: Cambridge University, 1998 (1781/1787).
- . *Groundwork of the Metaphysics of Morals*. 2nd ed. Translated by Mary Gregor and Jens Timmerman. Cambridge: Cambridge University Press, 2012 (1785).
- . *Religion Within the Limits of Reason Alone*. 2nd ed. Translated by Theodore M. Greene and Hoyt H. Hudon. New York: Harper Torchbooks, 1960 (1793).
- . *The Metaphysics of Morals*. Translated by Mary Gregor. Cambridge: Cambridge University Press, 1991 (1797).
- Korsgaard, Christine M. "Morality as Freedom." In *Kant's Practical Philosophy Reconsidered: Papers presented at the Seventh Jerusalem Philosophical Encounter, December 1986*, edited by Yirmiyahu Yovel. 1986.
- Pearce, Kenneth L. "Foundational Grounding and the Argument from Contingency." *Oxford Studies in Philosophy of Religion* 8 (2017).
- Sussman, David. "Perversity of the Heart." *The Philosophical Review* 114, no. 2 (2005): 153–177.