

Transformation and Personal Identity in Kant

Kant's account of transcendental freedom, particularly as developed in the *Groundwork* and in the first and second *Critiques*, seems to imply a radical discontinuity between an agent who acts at time t_1 and that same who acts at time t_2 . Transcendental freedom, it will be remembered, is the "power of absolutely beginning a state, and therefore also of absolutely beginning a series of consequences of that state."¹ Insofar as an action is transcendently free, it is an absolute beginning and is as such in no wise grounded in the prior state of the agent. This means that the two states, that of an agent before the initiation of action and that of the agent initiating the action, cannot be connected in a law-like manner. The positing of transcendental freedom would thereby seem to vitiate the unity of experience. These consequences are further reaching still when applied to the practical sphere: if we assume transcendental freedom it seems that we are not able to connect an action with an agent's character, that is, we are not able to understand how an action flows from the character of the agent, since assuming transcendental freedom, there is no way of connecting the action with the state of the agent preceding the action.

These issues have far-reaching consequences for how we are to understand moral development and character formation. If we are to envision ourselves as completely free at the moment of moral choice, the question of the cultivation of the soul can easily become one that is indifferent to us. F. D. Schleiermacher, for instance, charged that the idea of transcendental freedom is a dangerous one; it lulls people into unconcern regarding the formation of character and the care of the soul. It deludes them into thinking that at a moment's notice they can reverse the effects of a life ill-lived and all of a sudden start doing the right thing. As such, the whole train of states of the soul which must precede an individual's ability to act virtuously, in particular in situations requiring a great deal of moral fortitude, would simply be

ignored. The idea that at the moment of moral choice we are no more determined to choose one way than to choose another seems to imply that our past actions do not affect our present ones.²

This paper will explore how Kant's development of the idea of the disposition as developed in the *Religion Within the Bounds of Reason Alone* copes with some of these difficulties. I will argue that through it Kant is able to take into account the significance of character formation and the importance of instilling virtue while at the same time preserving transcendental freedom. Good and bad deeds do not occur in a vacuum; underlying them is the disposition, the ground of good and bad maxims. As the underlying maxim grounding all other maxims, the disposition serves as a principle unifying all actions of the self flowing from rule governed behavior. As such, the disposition is the seat of a person's character. While it connects an individual's various maxims made over time with one another by grounding them in an underlying maxim, it still preserves transcendental freedom since the disposition is itself freely chosen.

The paper is divided into two parts. In the first part I analyze Kant's understanding of the disposition and discuss the ways in which it allows us to understand a person's transcendently free actions in terms of that person's character. In order to analyze the deep structure of what a disposition is and what it entails for Kant's understanding of character, I compare Kant's view of personal identity with Parfit's Humean understanding of it.

In the second part of the paper I discuss Kant's resources for understanding the Socratic injunction to care for the soul in light of his concept of the disposition. Kant's critics have accused him of an inability to give a coherent account of how the phenomenal development of virtue is possible given his account of transcendental freedom. How is it possible to care for the soul if, in accordance with the doctrine of transcendental freedom, our present state does not determine our future actions?

Concern with one's character makes sense only if past states can influence future ones and the actions that flow from them.

I. Personal Identity and the Disposition

Kant introduces the concept of the disposition or *Gesinnung* in the context of his discussion of radical evil in the *Religion*. The disposition is the ground of all an individual's maxim making, although it is itself freely acquired. As such, the disposition serves as a bridge linking the maxims an individual adopts at time t_1 with the maxims adopted at time t_2 .

To have a good or an evil disposition as an inborn natural constitution does not here mean that it has not been acquired by the man who harbors it, that he is not author of it, but rather, that it has not been acquired in time (that he has always been good, or evil, from his youth up). The disposition, i.e., the ultimate subjective ground of the adoption of maxims, can be one only and applies universally to the whole use of freedom. Yet this disposition itself must have been adopted by free choice, for otherwise it could not be imputed. But the subjective ground or cause of this adoption cannot further be known (though it is inevitable that we should inquire into it), since otherwise still another maxim would have to be adduced in which this disposition must have been incorporated, a maxim which itself in turn must have its ground. Since, therefore, we are unable to derive this disposition, or rather its ultimate ground, from any original act of the will in time, we call it a property of the will which belongs to it by nature (although actually the disposition is grounded in freedom).³

Insofar as the disposition is the ultimate subjective ground of maxim-making, it functions much in the same way as a person's character. Through it we can understand continuities in a person's decisions and actions. When Kant says that the disposition is a "property of the will which belongs to it by nature" he means that the disposition is that which accounts for all the maxims that an agent makes in time. However, it is important to understand the relationship between the disposition and these lower-level maxims. A person's disposition does not function as a pre-existing condition that determines an agent to adopt a maxim—much in the same way that a complex of efficient causes would be sufficient to produce a given effect. It functions, rather, as the reason or principle that accounts for why a given maxim was chosen. This reason or principle in accordance with which we choose lower level maxims is itself freely chosen and is hence something acquired; it is something for which we are responsible.

The disposition is an agent's "highest" maxim; as such, it cannot be accounted for by invoking some higher-level maxim. For instance, if an individual makes it her maxim to lie when she finds herself in difficulty, we can always inquire into the ground of this maxim, that is, the reason for her adoption of it. Precisely insofar as we think of this maxim as having been freely adopted, and thus not adopted as the result of some natural impulse, the only explanation that can be given for the adoption of this maxim m_1 is another maxim m_2 , by virtue of which we can understand how it was that m_1 could have posed as an incentive to the will. In other words, once we posit freedom, the only reason that can be given for a person's having chosen to act on the maxim to lie when in difficulty is another maxim, one which accounts for the fact that this maxim was deemed worth acting on. But if we inquire into the ground of our adoption of m_2 , the answer can only lie in another maxim, m_3 . If we continue in our inquiry into the ground for the adoption of each of these maxims, we should be led back into an endless regress. Hence Kant notes

that the ultimate subjective ground of the adoption of moral maxims is inscrutable is indeed already evident from this, that since this adoption is free, its ground (why, for example, I have chosen an evil and not a good maxim) must not be sought in any natural impulse, but always again in a maxim. Now since this maxim also must have its ground, and since apart from maxims no determining ground of free choice can or ought to be adduced, we are referred back endlessly in the series of subjective determining grounds, without ever being able to reach the ultimate ground (R 17-18, note).

Reason, however, demands that we posit a kind of unconditioned condition in understanding the determination of our will. In other words, in explaining our adoption of maxims to ourselves we must reach a point where inquiry stops, and this point is a supreme or ultimate maxim which accounts for how all other maxims might pose as incentives to the will. Given that the adoption of this ultimate or supreme maxim is grounded in freedom, and no further reasons can be adduced for its adoption (otherwise it would be conditioned, and not a supreme or ultimate maxim) its adoption is inscrutable to us and constitutes the very mystery of freedom. This ultimate or supreme maxim, equivalent to the most general rule according to which an agent's more specific maxims are determined, conditions an agent's adoption of specific maxims, and constitutes an agent's fundamental character or disposition. According to Kant, this fundamental disposition cannot be accounted for by "any original act of the will in time" precisely because any such act occurring in time must be accounted for by another maxim, one which explains how such a maxim was able to pose as an incentive to the will. Kant's conclusion that this fundamental disposition is not something acquired in time is one arrived at negatively: given that any act occurring in time cannot be thought of as unconditioned (since it belongs to the phenomenal realm, and as such must be

thought to belong to the series of cause and effect), the unconditioned condition of all acts of the will in time cannot itself be thought of as occurring in time.

The disposition is, in essence, the fundamental choice that an individual makes, having a determining influence on all other choices.⁴ In the context of the *Religion*, it is the fundamental choice one makes as to which incentive will be valued over others. According to Kant, the human being must adopt *both* the moral incentive, and, since s/he is a sensuously conditioned being, the incentive of self love into her maxim. Since both are adopted into this maxim, either the satisfaction of one's self love will be made the condition of one's following the moral law, or action in accordance with the moral law will be made the condition of satisfying one's empirically conditioned desires. The distinction between a good and evil individual "cannot lie in the difference in the incentives which they adopt into their maxim (not in the content of the maxim), but rather must depend upon *subordination* (the form of the maxim) *i.e., which of the two incentives he makes the condition of the other.* (R 31). The evil individual is evil in that he "reverses the moral order of the incentives when he adopts them into his maxim," that is, he "makes the incentive of self-love and its inclinations the condition of obedience to the moral law. . . ." (R. 31-32). The choice of how these incentives are to stand in relation to one another constitutes the individual's fundamental disposition. This fundamental choice dictates the choices an individual will make in given circumstances, that is, it determines what will be valued in a particular situation and the agent's understanding of the necessary course of action to achieve it. Hence the fundamental disposition acts as a kind of filter, or lens through which the values assigned to things or situations is already determined beforehand; for instance, the individual that has chosen an evil disposition will value the fulfillment of his desires (whatever they may be) precisely because they are *his* desires over fulfillment of the moral law. This act of valuing, however, is such that the individual often does not understand him/herself as actively engaged in it, but

rather, already *finds* desires connected with self love as more valuable than the incentive posed by the moral law. This is one way in which the fundamental disposition can determine particular acts in time—it can function as a kind of filter *through* which particular situations are assessed.

In order to come to a better understanding of Kant's resources for making sense of character—the ground of a person's actions through time—we also need to explore the relation of the disposition to the transcendental unity of consciousness. While it might be objected that the latter belongs exclusively to Kant's theoretical philosophy and does not have a place in the practical sphere, reflection on the conditions necessary for rule governed behavior will show that in order to relate desires to one another in order to rank them—an operation indispensable to acting on principle—these desires must be apprehended as the desires of one consciousness. Acting on principle presupposes that in order to achieve a desired goal, means adequate to the end in question will be adopted. But this also means that fleeting desires, and perhaps other desires as well, must be deemed not worth acting on, since to adopt the means necessary to achieve a desired end often means foregoing the satisfaction of other desires. For instance, in order to achieve the goal of health, other desires, such as the desire to eat as many sweets as one pleases, must be ignored. To act on principle already presupposes that the agent has some awareness of a variety of his or her desires and has ranked them, foregoing those that are less important and that conflict with his/her more significant goals. Hence the transcendental unity of consciousness must be presupposed if we are to make sense not only of the possibility of rule governed behavior, but of the adoption of a fundamental disposition as well, since in order for the two fundamental incentives to be compared with one another and ranked, they must be the desires of one consciousness.

At this point it is instructive to ask: what kind of unity of the self does the having of a disposition require? What are Kant's resources for making sense of the unity of the self from a practical standpoint? This is important since the notion of character generally implies the idea of a subject existing through time; particular acts are grounded in the make-up of this subject. One way of beginning to answer these questions is to compare Kant's understanding of the practical unity of the self with Parfit's understanding of the self. On Parfit's view—he calls it the reductionist view—personal identity just consists in physical and psychological continuity. Parfit elaborates this reductionist view through the following two claims:

- 1) the fact of a person's identity over time just consists of the holding of certain more particular facts
and
- 2) these facts can be described without either presupposing the identity of the person, or explicitly claiming that the experiences in this person's life are had by this person, or even explicitly claiming that this person exists.⁵

On this view there is no *subject* of which thoughts and experiences are determinations; no subject that might supply continuity to the self throughout its changes. The view closely follows Hume's "bundle" theory, where the self is understood simply in terms of the various thoughts, desires, and experiences that are the denizens of a person's psyche. As Hume noted, when I introspect these objects of consciousness are all that I encounter; I can never come across the subject that has these thoughts and desires.⁶ Given that no subject is to be found, psychological continuity must be characterized in terms of causal relations between particular facts. For instance, one thought may trigger another, or the satisfaction or frustration of a desire may lead to another. The only possible criteria of continuity between an individual at time t_1 and that same individual at time t_2 (20 years later)

would be the number of facts of consciousness that have remained the same, or the kinds of causal connections holding between the series of the individual's thoughts and desires or states of consciousness. On Parfit's view, for instance, what matters is what he calls relation R: psychological connectedness or unity with the right kind of cause.⁷

On this view, there is no function of "transcendental unity" bringing these thoughts, desires, and experiences into the unity of one consciousness and hence into relation with one another. Now it is no doubt true that in his first *Critique* Kant took over some of Hume's fundamental insights when he argued against the possibility of theoretical knowledge of a substantial soul. From the *function* of the transcendental unity of the self we are not warranted in arguing to a *substance* that fulfills such a function. Nevertheless, Kant's understanding of the self and of personal identity differed significantly from Hume's theory precisely in that it did assume such a transcendental unity of the self as a condition of the possibility of experience.

The groundwork for Kant's theory of the self is already laid in his subjective deduction in the first *Critique*. There we find that the transcendental unity of apperception makes possible the unification of both synchronic and diachronic elements, and that both are necessary if experience is to be possible at all. Hence the synthesis of apprehension through which discrete impressions are run through and gathered together is intrinsically tied to the synthesis of reproduction in which representations must be actively reproduced from one moment to the next. In order to run through and gather together discrete impressions I must be able to reproduce past impressions that I have already run through so that I can associate them with what is presently before my mind. Kant's famous example concerning counting is to the point here:

For experience as such necessarily presupposes the reproducibility of appearances. When I seek to draw a line in thought, or to think of the time from one noon to another, or even to represent to myself some particular number, obviously the various manifold representations that are involved must be apprehended by me in thought one after the other. But if I were always to drop out of thought the preceding representations (the first parts of the line, the antecedent parts of the time period, or the units in the order represented), and did not reproduce them while advancing to those that follow, a complete representation would never be obtained: none of the above-mentioned thoughts, not even the purest and most elementary representations of space and time, could arise.

The synthesis of apprehension is thus inseparably bound up with the synthesis of reproduction (*KRV*, A 102).

Furthermore I must be able to recognize these past impressions as *my own*; there are analytic connections between the ability to synthesize the manifold and the thoroughgoing identity of self-consciousness. As Kant notes,

In other words, only in so far as I can grasp the manifold of the representations in one consciousness, do I call them one and all mine. For otherwise I should have as many-coloured and diverse a self as I have representations of which I am conscious to myself. . . . This principle of the necessary unity of apperception is itself, indeed, an identical, and therefore analytic, proposition; nevertheless it reveals the necessity of a synthesis of the manifold given in intuition, without which the thoroughgoing identity of self-consciousness cannot be thought (*KRV*, B 134).

On Kant's view we cannot simply speak of the *facts* of consciousness as if these facts could exist in isolation from one another and from the thoroughgoing identity of self-consciousness itself. These facts must be relatable to one another, and if they are relatable to one another it can only be in virtue of the identity of self-consciousness which first represents itself in relation to one state of affairs (for instance as knowing or desiring x) and then in relation to another. This cannot be stressed enough. For Kant the very possibility of experience involves a transcendental unity of apperception that makes both synchronic *and diachronic* connections.

In this regard it may be objected that Kant himself admitted that the identity of the consciousness of myself at different times does not prove the numerical identity of my subject. States of consciousness, for instance, might be passed on baton-like from one self to another. In such a case different *substances* exchange properties or determinations:

we can conceive a whole series of substances of which the first transmits its state together with its consciousness to the second, the second its own state with that of the preceding substance to the third, and this in turn the states of all the preceding substances together with its own consciousness and with their consciousness to another. The last substance would then be conscious of all the states of the previously changed substances, as being its own states, because they would have been transferred to it together with the consciousness of them. And yet it would not have been one and the same person in all these states (*KRV*, A 364).

In such a scenario the self-same individual would be conscious of different states connected with the past, but there would be no guarantee that *that* very same

individual was the one that actually had those states of consciousness at some point in the past. At most, what the formal unity of consciousness allows us to infer is that “in the whole time in which I am conscious of myself, I am conscious of this time as belonging to the unity of myself” (*KRV*, A362). Hence Kant concludes that this simply amounts to saying that this time is in me. It is, however, important to note that here we do not simply have *facts* of consciousness, since the formal unity of consciousness holds them together, making it possible to set them in relation to one another.

Moreover, it is important to take into account the *context* in which Kant makes these remarks. The issue at stake in the third paralogism of transcendental psychology is how we can ascertain the numerical identity of the self *through experience*. In order to do so we need to find “that permanent element in the appearance to which as subject everything else is related as determination” (*KRV*, A362). However, the transcendental unity of consciousness can never be an object of experience, and as such can never be identified with a permanent element in appearance. It is for this reason that we cannot rule out the possibility of states of consciousness being passed on baton-like from one self or subject to the next. Moreover, given that the problem here is how we can ascertain the numerical identity of the self *through experience*, whether there is a problem regarding personal identity depends upon one’s *standpoint*. Given that from the first person point of view—the standpoint of the agent—this whole time *is in me*, Kant concludes that “in my own consciousness. . . . identity of person is unfailingly met with” (*KRV*, A362). Establishment of personal identity crops up as problematic only from the standpoint of another person, that is, from a third person point of view:

But if I view myself from the standpoint of another person (as object of outer intuition), it is this outer observer who first represents *me in time*, for in the apperception *time* is represented, strictly speaking, only

in me. Although he admits, therefore, the ‘I’ which accompanies, and indeed with complete identity, all representations at all times in *my* consciousness, he will draw no inference from this to the objective permanence of myself (*KRV*, A362-363).

The problem of character and the care of the soul only arises, however, from a practical standpoint, one which first and foremost presupposes a first person transcendental point of view. Here then we not only have the identity of a person “unfailingly met with,” but we also must take into account 1) the function of the transcendental self in relating a person’s representations, experiences and desires to one another, as well as 2) the relation of this freely acting transcendental self to its actions in time. For it is only in the context of a life, or at least of a long period of time, that it makes sense to speak of a person’s character, its development, and the relation of this character to discrete actions.

An individual is not simply a bundle of disparate experiences, desires, representations, in short, of facts of consciousness. For these desires, experiences and representations are unified and relatable to one another insofar as they are those of a *single* self. And this being a single self is something *more* than Parfit’s reductionist view of having a brain and a body. It rather involves the transcendental activities of relating representations to one another so as to make experience possible on the one hand and of the structuring of desire so as to make action possible on the other. If a person’s representations, experiences and desires are to influence one another they must be the representations, experiences, and desires of a single transcendental consciousness.

Now while such a transcendental consciousness can never become an *object* of experience, we are conscious of its activities and functions when we think and act. The first person point of view involved in action involves us in what Christine Korsgaard has dubbed the authorial point of view.⁸ From this point of view we view

actions and choices as essentially our own rather than as things that simply happen to us. A person is not simply a nexus of experiences and desires moved to act on whatever desire happens to be the strongest one. The agent evaluates and gives structure to his/her desires; she weighs their relative worth in light of her picture of a life worth living. Even if the individual decides not to evaluate her desires, but simply decides to act on her strongest desire, this is in itself an evaluative choice—the choice to live the unexamined life and to be swayed by the moment's desire. Henry Allison makes a similar point in his defense of Kant's concept of transcendental freedom: an agent does not act on desires simply because she has them, even when s/he acts heteronomously. In all cases, a desire must be *incorporated* into an agent's maxim, which means that the simple having of a desire is not a sufficient reason for acting upon it; an agent must also deem it a desire *worth* acting upon. This act of incorporating a desire into one's maxim cannot be thought of as itself the effect of some other, higher order desire determined in us by causes lying outside the will. Rather it must originate with the will itself, and this means that it must be an act of transcendental freedom.⁹ As argued above, once freedom is assumed, when we seek to understand *why* a given maxim was acted upon, the answer can only lie in some other higher order maxim that was itself freely adopted. The highest order maxim is the disposition, which grounds all lower level maxims adopted in time. It is through the disposition that we can relate a maxim made at time t_1 with a maxim made at time t_2 ; moreover, through the disposition we can understand continuities in a person's character. Now the adoption of a disposition is an act of the free transcendental consciousness, and so is likewise the act of evaluating, ranking and ordering desires in relation to one another in light of the fundamental disposition.

If the notion of character is to make sense, we cannot speak of a bundle of desires and experiences loosely related to one another. They must be related to one

another in virtue of one consciousness that is able to survey these experiences and desires together, to evaluate them and to rank them. Similarly, if we are to speak about character we must also be able to understand how actions done at two different periods of time are related to one another. The question of personal identity through time thus becomes an issue crucial to the problem of moral development and character formation. If I cannot assume that some kind of identity holds between myself today and myself five years from now, but rather I must consider the time throughout which I exist as containing a series of selves loosely related to one another, the whole notion of goal oriented action loses its coherence. Similarly the whole idea of progress in virtue and of the care of the soul presupposes that the person I am now is one and the same as the person who I will be five years from now. If I am to progress in virtue, *I* must be able to undergo change as I effect change in myself; if there is no ‘I’ that undergoes change but simply a loose collection of selves related to one another in terms of causal relations and degrees of similarities, it is not clear why I should be concerned about the state of these future selves.

Kant’s solution to the problem is that from a first person transcendental point of view time is *in me*, that is, “in the whole time in which I am conscious of myself, I am conscious of this time as belonging to the unity of myself” (*KRV*, A362). In fact, as a form of inner sense time is generated by my act of attending first to one representation and then to another. From a practical perspective this means that insofar as I project myself into the future, I am conscious of these future experiences as my own and hence of the time in which they occur as belonging to the unity of my consciousness. Moreover, our present projects take time and require us to project ourselves into the future, i.e., we do now what it is that will be necessary to achieve a desired result later. This means that from a practical standpoint *I must* identify with my future self if my present actions are to have any meaning; as Korsgaard puts it, “to

the extent that you regulate your choices by identifying yourself as the one who is implementing something like a particular plan of life, you need to identify with your future in order to be *what you are even now*.¹⁰ All of this is important, in that it shows that Parfit's attempt to derive ethical consequences concerning how we should regard our future selves (as not strictly identical with our present self) is mistaken.

Contra Parfit, two things have been shown. First, that the self is not simply a bundle of facts of consciousness since experience is only possible if we assume the transcendental unity of consciousness. Second, that from a first person practical point of view I can and must assume personal identity throughout time; from a practical standpoint, it is impossible not to identify with my future self.

II. Transformation, Personal Identity, and the Development of Virtue

If Kant is to have a viable account of how it is possible for an individual to cultivate virtue, he needs to have the following elements in place. First, he needs to show how we can understand the identity of an individual through time. Second, he needs to provide us with an account of how it is that actions flow from a person's character, that is, how it is possible that the state of the individual before initiating an action is connected with the initiation of that act itself. If this link were not established, we could not connect a person's intentions and desires (elements of a person's character) with what that person actually does at a given moment. These two issues have been discussed above. Third, he needs to provide us with some account of what exactly it means for an individual to *grow* in virtue.

At this point it is instructive to compare Kant's understanding of virtue with that of Aristotle. For Aristotle, to acquire virtue is a matter of habituating the self to act virtuously until virtue becomes a matter of second nature. Hence, one begins with concrete actions, and these actions eventually have an effect on

character; it is as if acting in a certain way etches certain pathways in the psyche and *informs* it. Particular actions thus become constitutive of character. For Kant, on the other hand, an action only has moral significance insofar as it can be understood as action in accordance with a rule. Furthermore, an action has moral worth only if it is done for the sake of duty; it is not enough for it to be merely in accordance with duty. Hence to believe that given actions, regardless of their motivation, can become habitual and can thus serve to build up a person's moral character is mistaken. To believe that one can go about reforming the fundamental ground of the will through singular actions is a confusion of cause and effect. Particular actions, being the fruit of a will that is itself either good or evil, are always merely its concrete expressions, and it is a misunderstanding to believe that by merely changing *particular actions*, the fundamental maxim of the will itself will be changed.¹¹

For Kant, then, it seems that the only way we can connect an action at time t_1 with an action at time t_2 is through the disposition. The two actions are not *directly* related to one another. Rather, they are related to each other only in virtue of some third thing to which each is related, namely the disposition itself. Hence it seems that insofar as given actions are merely the concrete *expressions* of the fundamental disposition, there is no way that a particular action in time can influence the seat of character. Given such an account, it becomes very difficult to understand the possibility of the phenomenal development of virtue where concrete actions serve to constitute a person's character; instead, it seems that all actions are simply the expressions of a fundamental disposition, itself timelessly chosen.¹² Allen Wood notes that

timeless agency . . . makes nonsense of the goal of moral improvement or moral progress. . . . On Kant's theory, only our external actions occur in time; our freedom, to which alone true moral worthiness or unworthiness pertains, is timeless, and hence incapable of changing for

better or worse. . . . there is no place in Kant's theory for the idea that I may resist some passion or inclination of mine tomorrow by struggling with it today, and by striving throughout the day to purify my motives and fortify myself for the crucial hour of decision.¹³

Kant's view seems completely opposed to the Aristotelian one, where particular actions are understood to have an effect on the seat of moral character.

However, the picture that Kant offers us in the *Religion* is much more complicated than this. Key to understanding his account of progress in virtue is his characterization of the propensity to evil. Such a propensity is the *effect*, and not the *ground*, of our having adopted a fundamentally evil maxim.¹⁴ The adoption of such a fundamentally evil maxim, in which the satisfaction of the lower faculty of desire is made the condition of obedience to the moral law (R 32), has effects that far outlast the time during which such a maxim is operative. Even after the individual, "reverses, by a single unchangeable decision, that highest ground of his maxims whereby he was an evil man (and thus puts on the new man)" (R43) s/he will have to deal with the propensity to evil. Hence the person can be considered good "only in continuous labor and growth" (R 43).

Kant defines a propensity as a 'predisposition to crave a delight which, when once experienced, arouses in the subject an inclination to it' (R 25). Hence a propensity is linked with the frailty of our human nature, which is partly rational, partly animal. Once an evil disposition has been operative in our lives, it adversely affects the structure of our desiring *even after we have resolved to become better persons*. The propensity to evil functions much in the same way as an addiction: one can, through a free act, incapacitate the freedom of the will through the affection of one's sensibilities; henceforth one craves a particular substance. Analogously, once the incentive of self-love has been valued over the moral law, there will be a predisposition to want to put oneself and one's subjective interests above duty.

The dynamic of how a propensity to evil can coexist with a fundamentally good disposition can be clarified through a distinction between first and second order desires. First order desires are those desires I simply happen to have; they are immediately *given*. Second order desires concern the person I would like to be, and hence have to do with the kinds of desires I would like to have.¹⁵ For instance, I may want to be the kind of person that does not feel envy or jealousy (a second order desire), yet may still find myself coveting my neighbor's good fortune. The fact that I really do want to be the kind of person that takes joy in another's good fortune would reflect my having adopted a fundamentally good maxim; this however, could in principle coexist with negative feelings and desires with which I would rather not identify, but which I nonetheless find myself having. The clearest example of the coexistence of the propensity to evil with a fundamentally good disposition can be found in Kant's discussion of the first degree of radical evil:

. . . the frailty of human nature is expressed even in the complaint of the Apostle, "What I would, that I do not!" In other words, I adopt the good (the law) into the maxim of my will, but this good, which objectively, in its ideal conception (*in thesi*), is an irresistible incentive, is subjectively (*in hypothesis*,) when the maxim is to be followed, the weaker (in comparison with inclination) (R 25).

In the case of the first degree of radical evil, not only is the individual torn by a discrepancy between her first and second order desires, she also finds that the desires she simply finds herself having, and with which she would rather not identify, (her first order desires) frequently overpower her. These first order desires are the left over effects of once having adopted a fundamentally evil maxim.

That a basically good individual can be incapacitated through the propensity to evil is a fact crucial to understanding how progress in virtue is possible over time.

The fact that an individual can make progress in virtue already presupposes that s/he is committed to a good course of life without, however, already having achieved perfection. One way of explicating this would be to say that the individual's first order desires were not in line with her second order desires. Progress in virtue would thus consist in bringing first order desires in line with second order desires. This would imply, however, that our first order desires can be strengthened or weakened, that they can be affected and brought into line with our second order desires. Insofar as this is true, particular actions *can* have an affect on the desires that we simply find ourselves having, i.e., on desires that are often associated with the lower faculty of desire and hence with our *receptivity*.¹⁶ For instance, whether or not I chose to start to smoke today will have an affect on the sorts of desires I will have tomorrow; the desires I have tomorrow will in part depend on what it is I chose to invest my energies in today. Hence temporal actions are not simply the expressions of a timelessly chosen disposition. *Which* desires I ultimately incorporate into my maxim and act on today will depend upon what I did yesterday and the kind of projects I invested myself in. This means that I can affect myself (in accordance with my second order desires) so as to channel and direct my first order desires.

Numerous passages show that Kant was well aware of this. For instance, in the *Anthropology* Kant notes that "Sensitivity is a *power* and *strength* by which we grant or refuse permission for the state of pleasure or displeasure to enter our mind, so that it implies a choice."¹⁷ Insofar as one's desires are often connected with one's emotional state, the affection of one's emotions will also have an effect on one's desires. And in his *Lectures on Ethics* Kant claims, "The command to love our neighbor applies within limits both to love from obligation [practical love] and love from inclination [pathological love]. For if I love others from inclination, I acquire in the course of time a taste for it, and my love, originally duty born, becomes an inclination."¹⁸

This bringing of my first order desires into line with my (moral) second order desires constitutes growth in virtue. This analysis comes closer to the Aristotelian picture where virtues are acquired habits, although there are of course significant differences between Kant and Aristotle. In both the Kantian and Aristotelian pictures desires can be channeled through concrete temporal actions. For Kant, however, in order for this to count as growth in *moral* virtue, the channeling of my first order desires must be the result of conscious self-affection in accordance with moral second order desires; growth in virtue cannot simply be a question of habituation.

While Kant acknowledges the possibility of growth in virtue, he insists that “if a man is to become not merely *legally*, but *morally*, a good man . . ., *this* cannot be brought about through gradual reformation so long as the basis of the maxims remains impure, but must be effected through a *revolution* in the man’s disposition” (R. 43). Growth in virtue is indeed possible, but it is something that occurs only after the revolution in the disposition has occurred.

Kant’s account of personal identity encounters its severest difficulties in regard to the problems associated with a revolution in the disposition. If the disposition functions as the underlying unifying ground of rule governed behavior, how are we to account for the change from one disposition to the other? If all that unifies maxim-making is the disposition, does it not then seem as if there is no connection between the individual before and after conversion when s/he, through a fundamental decision, alters his or her disposition? Is there any way that we can understand personal identity throughout this change? It is significant that in his discussion Kant adopts the Pauline image of the “death of the old man” and the “birth of the new.”¹⁹ Moreover, he continually notes that the individual that has exchanged an evil disposition for a good one is “morally another person.”²⁰ Given that the disposition is the only viable candidate available to explain how we are to

connect an agent's actions at two different times with one another, there is much to be said for this kind of language. After all, as noted above, it is illegitimate to think of the individual as a kind of *substance* in which thoughts and character traits inhere and which endures throughout changes in its thoughts and objects of experiences.

If an individual cannot be this kind of substantial entity, then neither can a character. At best we can understand a character as a set of concerns, projects and desires had by an individual.²¹ Some of these concerns, projects, and desires are “nested” in projects and concerns of a higher order; for instance, one often undertakes projects for the sake of others. At the root will be a fundamental concern in terms of which other desires and projects can be understood. Hence a crucial component of character is just such a nexus of desires that stand in a hierarchical relation to one another: for example, the fulfillment of particular desire may be important only because of its relation to other, more fundamental desires that an agent has deemed worth pursuing. The disposition functions much in the same way as this root concern; once it is changed other desires must follow suit if the individual is not to be a divided self. Full integration of the personality takes time, since the person is often not aware that her desires in fact conflict with one another; experience is necessary for an individual to discover the real nature and significance of many of the things that s/he desires and how these may be incompatible with his or her fundamental life project (for the good individual this would amount to a valuing of action for the sake of the moral law above all else). This is another way of understanding how the having of an evil disposition leaves a “residue” even after it has been exchanged for a good one; one often continues to value elements expressive of the old self's fundamental project without understanding how they are connected with it. For example, an individual may have undergone a genuine moral conversion while still having commitments (for instance making a great deal of money) that might conflict with the demands of social justice. It is often not clear to an

individual how moral conversion has implications for all of a person's commitments: personal, social, and political. A person that has undergone such a conversion will no doubt be open to growth in understanding and a broadening of his or her horizons in all of these areas. This broadening, however, often takes time and education.

Once the disposition, or root concern has been changed, there is a fundamental sense in which the individual is a different person. How a change in disposition is possible is a mystery; it is connected with the problem of the transcendently free self that chooses. Kant himself notes that “. . . since the lapse from good into evil (when one remembers that this originates in freedom) is no more comprehensible than the re-ascent from evil to good, the possibility of this last cannot be impugned” (R 40). As noted above, the transcendental self does the work of relating a person's thoughts, desires, and experiences to one another. Because it is never an object of experience, very little else can be said about it aside from the way it functions. *Why* it would choose to make the self its fundamental project as opposed to action in accordance with the moral law (radical evil), or how it is possible that it should change from evil to good remains shrouded in mystery. Given the fact that little can be said about the transcendently free self, there seem to be very few resources available to Kant to provide an account of personal identity throughout this change.

There are, however, a few things that Kant *can* say regarding the problem. First, while there is no explanation for the descent into evil, there is an important sense in which it is only when the self acts in accordance with the moral incentive that it is at one with itself. To choose the incentive of self love over the moral incentive (which can never be eradicated) means, for Kant, to act on desires one simply *happens* to have and hence to allow one's will to be directed by causes lying outside of it.²² The individual that chooses to act on whatever desire s/he happens to have regardless of the directives of the moral law acts from the incentive of self-love.

One's empirically given desires (the desires one happens to have) however, are caused. Since every cause is also an effect, these causes can eventually be traced to causes lying outside the self. To prioritize the incentive of self-love over that of the moral law is thus, ironically, to identify with something that is not really oneself. Regardless of the descent into evil, the true identity of the self lies with *Wille*, the ability of the will to be a law unto itself. This ability is never lost by the person; as Kant notes the ground of radical evil cannot "be placed in the corruption of the morally legislative reason—as if reason could destroy the authority of the very law which is its own, or deny the obligation arising therefrom; this is absolutely impossible" (R 30).²³

Second, we have already discussed the phenomenon of radical evil—the volitional residue left over even after one has exchanged an evil maxim for a good one. That past desires can affect future ones in this way is evidence for the unity of the self through time. Because it is the *same* self that once chose an evil maxim that now chooses a good one, the effects of that past choice must be undone. The nexus of a person's desires are related to one another by the rational agent. Because they are so related, choosing to act on a given desire often affects one's other desires. Moreover, the adoption of an evil fundamental maxim affects the very structuring of desire and hence the whole way that an individual values and understands the world. Undoing these effects takes time and growth in virtue. Hence in a remarkable passage worth citing at length Kant asks whether the person who has entered into goodness should count the evils that occur to her in this life as punishment:

Yes, but only in the quality of the man whom he is continually putting off. Everything (and this comprises all the miseries and ills of life in general) that would be due him as punishment in that quality (of the old man) he gladly takes upon himself in his quality of new man simply for the sake of the good. So far as he is a new man, consequently,

these sufferings are not ascribed to him as punishment at all. The use of the term “punishment” signifies merely that, in his quality of new man, he now willingly takes upon himself, as so many opportunities for the testing and exercising of his disposition to goodness, all the ills and miseries that assail him, which the old man would have had to regard as punishments and which he too, so far as he is still in the process of becoming dead to the old man, accepts as such. This punishment, indeed, is simultaneously the effect and also the cause of such moral activity and consequently of that contentment and *moral happiness* which consists of a consciousness of progress in goodness (and this is one and the same act as forsaking evil). While possessed of the old disposition, on the other hand, he would not only have had to count the very same ills as punishments but he would also have had to *feel* them as such, since, even though they are regarded as mere ills, they are the direct opposite of what, in the form of *physical happiness*, an individual in this state of mind makes his sole objective (R 69).

Becoming dead to the old man is a *process* and not something that happens at once. It includes the thwarting of many of the desires that are merely empirically given and so connected with self-love and with the body. It is often through ills thwarting these desires that moral learning takes place and that the self is purified of the vestiges of desire connected with the old self.²⁴

It is important to note that when we speak of moral learning and of progress in virtue we speak of one *self* that undergoes a change. This is the self that evaluates, ranks, and connects desires to one another and that slowly comes to understand the meaning of the having of a good disposition for all of the self's projects. Through such progress the person achieves a fully rounded character having all the desires and projects appropriate to an individual with a good disposition.

The fact that one's having chosen a bad disposition in the past still has an effect on the present (even after one has reversed the fundamental disposition) is once more evidence for the unity of the self through time. No doubt the skeptic can still object that it is possible that all one's desires are passed baton-like from one temporal part of the self to the next, without there being a single unifying principle relating all of these desires and hence all of these temporal parts to one another. However, such a hypothesis does not do as good a job of explaining how it is possible that present projects and desires can be related to one another (and to future ones) as the notion that these projects and desires are related to one another through the transcendental activity of practical rationality. One's past choices have an effect on one's future desires because both are related to one another through the transcendental activity of consciousness. A past evil disposition can have effects spilling over into the present even when the fundamental maxim has been changed because complete integration of the personality—the achievement of coherence among one's desires—is a process that takes time.

For Kant the self that chooses to act heteronomously is always a divided self. It is divided because the moral incentive can never be expunged and even the most wicked of persons feels its tug. Moreover, as noted above, to prioritize incentives that are merely *given* (and thus caused by factors lying outside the will itself) is to value, and hence to identify with something that is not really oneself. Full integration of the personality can thereby only occur when a good disposition has been chosen and the individual learns to bring all of her desires in line with the demands of morality. This integration is the goal of the cultivation of character and the care of the soul.²⁵

ENDNOTES

¹ *Critique of Pure Reason*, trans. N. Kemp Smith (New York: St. Martin's Press, 1965), References to the *Critique of Pure Reason* are to the standard A and B paginations of the first and second editions and will henceforth be included in the body of the paper preceded by the letters KRV. In this case the reference would appear as KRV A445/B473.

² On this point, see my article "Schleiermacher on the Philosopher's Stone: the Shaping of Schleiermacher's Early Ethics by the Kantian Legacy," *Journal of Religion*, 79, April, 1999, 193-215. Schleiermacher develops the point in his essay *On Freedom*, trans. by Albert L. Blackwell, (Lewiston: Edwin Mellen Press, 1992), 100-101. The point is one that was already developed by Leibniz, Hume, and other compatibilists: an action must have a sufficient reason grounded in the prior states of an agent; to deny this is to deny the conditions under which the act could intelligibly be attributed to that agent. A condition of act attribution is that we should be able to relate an action to an agent and his or her character, that is, we must be able to understand how it flows from that character. It seems, however, that if transcendental freedom is assumed, no such connection between the action and the character of the agent can be made. Schleiermacher asks, "How can I be accountable for an action when we cannot determine the extent to which it belongs to my soul?" (100-101). Our ability to attribute the motive for an action to an agent depends on that actions' being explicable in terms of an agent's character. Failing such a condition, the actions "have no ground at all, not even immediately, and are based on chance" (101), which means they have nothing to do with the condition of the agent, that is, his or her psychological states and disposition. Schleiermacher concludes that this idea of "complete chance . . . certainly annuls morality more than anything else" (101).

³ Immanuel Kant, *Religion Within the Limits of Reason Alone*, trans. by Theodore M. Greene and Hoyt H. Hudson, (New York: Harper & Row, 1960), 20-21. Henceforth all references to the *Religion* will be cited in the body of the text indicated by an ‘R’ followed by the page number of the Greene and Hudson translation.

⁴ My understanding of how the disposition functions is directly at odds with the view put forward by G. Felicitas Munzel in her recent book *Kant’s Conception of Moral Character*, (Chicago: University of Chicago Press, 1999). There she notes that “through actual acts of choice making, the power of choice itself acquires the moral comportment (*Gesinnung, sittliche Haltung*) of the spirit of the maxims it has adopted in its acts of choosing. For maxims are not only principles of actions, but constitutive principles formative of character” (143). It seems clear from what Kant notes in the *Religion* that the disposition is the ground of actual acts of choice making and not the result of these choices.

⁵ Derek Parfit, *Reasons and Persons*, (Oxford: Oxford University Press, 1984), 210.

⁶ David Hume, *A Treatise on Human Nature*, edited by L. A. Selby-Bigge with revised text by P. H. Nidditch, (Oxford: Oxford University Press, 1978). So Hume, “When I turn my reflection upon *myself*, I can never perceive this *self* without one or more perceptions; nor can I ever perceive any thing but the perceptions. ‘Tis the composition of these, therefore, which forms the self” (634).

⁷ Parfit, *op. cit.*, 210.

⁸ Christine M. Korsgaard, “Personal Identity and the Unity of Agency: A Kantian Response to Parfit,” *Philosophy and Public Affairs*, 18, No. 2., 101-132

⁹ Henry Allison, *Kant’s Theory of Freedom*, (Cambridge: Cambridge University Press, 1990), in particular pages 40 and 51, although the theme is discussed throughout the book.

¹⁰ Korsgaard, *op. cit.*, 113-114.

¹¹ As Kant notes in the *Religion*, "... man's moral growth of necessity begins not in the improvement of his practices but rather in the transforming of his cast of mind and in the grounding of a character; though customarily man goes about the matter otherwise and fights against vices one by one, leaving undisturbed their common root (R 43-44).

¹² Allen Wood discusses the issues in his essay "Kant's Compatibilism," in *Self and Nature in Kant's Philosophy*, Allen W. Wood, ed. Ithaca and London: Cornell University Press, 1984, 73-101.

¹³ *Ibid.*, 97-98. Allen Wood has pointed out to me that on his view Kant's theory of timeless agency is simply the best theory Kant is able to come up with in rendering freedom consistent with natural causality; Kant is, however, not committed to it.

¹⁴ Gordon E. Michalson is certainly mistaken when he thinks of the propensity to evil as itself grounding our adoption of a fundamentally evil maxim. This misunderstanding runs throughout the whole of chapter two of his book and is well summarized by his statement that "the notion that we are evil by nature due to a natural propensity to adopt an evil disposition" is one of Kant's "key metaphysical points." 63. *Fallen Freedom: Kant on Radical Evil and Moral Regeneration*, (Cambridge: Cambridge University Press, 1990). Kant clearly notes in the *Religion* that "a propensity is distinguished from a predisposition by the fact that although it can indeed be innate, it *ought* not to be represented merely thus; for it can also be regarded as having been *acquired* (if it is good) or *brought* by man *upon himself* (if it is evil)" (R 24). The propensity to evil is itself the *result* of our having adopted a fundamentally evil maxim and its effects remain as a kind of residue (much like Augustine's concupiscence) even after the fundamental maxim of the will has been changed. For a further discussion of this point see my article "Kant on Grace: A Reply to his Critics," *Religious Studies* 33, 1997, 379-400.

¹⁵ The distinction is introduced by Harry Frankfurt in his essay entitled "Freedom of the Will and the Concept of a Person." It can be found in his book *The Importance of What we Care About*, (Cambridge: Cambridge University Press, 1988), 11-25.

¹⁶ For an good discussion of Kant's awareness that we can affect ourselves and so influence the course of our emotions and desires, see Nancy Sherman, "The Place of the Emotions in Kantian Morality," in *Identity, Character, Morality: Essays in Moral Psychology*, edited by Owen Flanagan and Amelie Oksenberg Rorty, (Cambridge: The MIT Press, 1990), pp. 149-170). As Sherman notes, for Kant "the fact that the emotions are states of being acted upon (ways of being affected, *pathe* in Greek), does not imply that they must be involuntary," (155).

¹⁷ Immanuel Kant, *Anthropology from a Pragmatic Point of View*. Trans. by M.J. Gregor (The Hague: Nijhoff, 1974), 104.

¹⁸ Immanuel Kant, *Lectures on Ethics*. Translated by L. Infield, New York: Harper & Row, 1963), 197.

¹⁹ One of the passages most notable for its use of Pauline imagery occurs in the second book of the *Religion*, where Kant attempts to make sense of the ideas of punishment and regeneration: "Now a change of heart is a departure from evil and an entrance into goodness, the laying off of the old man and the putting on of the new, since the man becomes dead unto sin (and therefore to all inclinations so far as they lead thereto) in order to become alive unto righteousness. . . . The coming forth from the corrupted into the good disposition is, in itself (as 'the death of the old man,' 'the crucifying of the flesh') a sacrifice and an entrance into a long train of life's ills" (R 68; cf. R. 43).

²⁰ For instance, at R. 67 and R. 68.

²¹ This is the way that Bernard Williams describes the having of a character in "Persons, Character, and Morality," in *The Identities of Persons*, edited by Amélie

Oksenberg Rorty, (Berkeley: University of California Press, 1976), 197-216. While Williams description of what constitutes character is highly useful, he is mistaken regarding what he calls “the Kantians’ omission of character” (210).

²² For a discussion of this point, see Andrews Reath in “Hedonism, Heteronomy, and Kant’s Principle of Happiness, *Pacific Philosophical Quarterly*, 70 (1989) 42-72.

²³ See also R 31.

²⁴ On this point, compare Hebrews 12.

²⁵ An earlier version of this paper was read in March, 1999, at a meeting of the North American Kant Society (with the American Society of Eighteenth Century Studies) in Milwaukee, Wisconsin. Research support was provided by a Purdue Research Foundation 1999 Summer Faculty Grant. I would also like to thank Manfred Baum, Franklin Mason, Fred Rauscher and Allan Wood for comments on earlier drafts.