

[*History of Philosophy Quarterly* 15, No. 3, July 1998: 349-367]

WEAKNESS INCORPORATED¹

Robert N. Johnson

U. of Missouri

Kant held that "an incentive can determine the will [*Willkür*] to action *only so far as the individual has incorporated it into his maxim*",² a view dubbed the "Incorporation Thesis" by Henry Allison (hereafter, "IT").³ Although many see IT as basic to Kant's views on agency, it also seems irreconcilable with the possibility of a kind of weakness, the kind exhibited by a person who acts on incentives that run contrary to principles she holds dear.⁴ The problem is this: According to IT, if an incentive determines the will of the weak person when she acts contrary to her principles, then it must be the case that she incorporated that incentive into her maxim. But that in turn means that she has made it her *principle* to act on the wayward incentive, and so is not, after all, exhibiting weakness in failing to follow her own principles, but at best simply dropping one principle in favor of another.⁵ So either the weak person *does not* incorporate the wayward incentive into her maxim and IT is false, or she *does* incorporate it and weakness is impossible.

I am going to argue that if we broaden our view of IT, we can see the weak as both incorporating wayward incentives into maxims, yet acting contrary to principles that they have adopted and hold dear. Of course, weakness of will is a complex and controversial subject on its own, as much or more so than Kant's

moral philosophy. Moreover, my discussion will draw on texts that raise thorny exegetical issues as well, for instance, texts in which Kant appeals at crucial points to mysterious "noumenal" choices made by a person's "intelligible" character. However, my concerns will not be primarily exegetical. Instead, my hope is that, without relying too heavily on controversial positions or texts, I can reconstruct some of Kant's views, found in several works, in a way that will illuminate what appear to many to be confused and implausible doctrines. So, although I will have to set aside important philosophical and exegetical issues, I think that nevertheless it will be profitable to consider some of these elements of Kant's moral psychology on their own, and I think there is enough textual support to see what follows as capturing much of what is important in his views.

1.

Kant famously rejects the view that actions are simply the effects of internal states, such as desires or incentives, whose origins are "alien" or external to agents.⁶ However, this rejection is only justifiable when restricted to *practical* contexts. In other contexts -- for instance, for the purposes of scientific study -- human behavior can be regarded as entirely causally determined.⁷ Hence, for Kant, reason is practical *for all practical purposes*, even if some other view might be justified for other purposes. The primary practical context is practical deliberation. So, regarding the context of deliberating over alternatives, Kant holds that an agent must regard these alternatives as *real* options, her choice of one not

being simply the effect of given desires or incentives, but rather the "spontaneous" product of her own practical deliberation.

As such, human actions are always grounded in what Kant calls "maxims". Although the meaning of "maxim" is controversial, and, as I will contend in a moment, ambiguous, the basic idea is clear enough.⁸ "Maxim" is Kant's term for an agent's practical principle or reason for acting. Maxims are thus more or less general principles of behaving in certain ways to gain certain ends given certain circumstances.⁹ For instance, my reason for riding a bike rather than driving to work might be to economize, for jogging regularly, that it will maintain my health. Hence, "I will ride my bike to economize" and "I will jog regularly to maintain my health" are roughly what Kant has in mind.

It is tempting to think of a "maxim" as simply a tidied-up answer to a "Why?" question regarding one's actions. But, of course, the answers we give to such questions are notoriously liable to be already more tidy than the truth, if not because we are self-serving, then because in order to answer at all we must impose order on things that often seem quite obscure to us. Nevertheless, our attempts to answer "Why?" questions do seem easily reconstructed in the form of a maxim, and it is not unreasonable to suppose that therefore this is how we view our own actions -- as grounded in the practical principles we have adopted, if not always those we find ourselves tempted to offer.

Rather than viewing incentives as brute causes of her future behavior, then, an agent must conceive of them as presenting demands to weigh, in the form of maxims, to further or realize (or not) certain ends. She must, inasmuch as she is deliberating, regard herself as engaged in a process the outcome of which will be a conception of an action in her circumstances that will produce some end, a conception that she must regard as the origin of her action. To "incorporate" an "incentive" into one's maxim is thus, in more plain language, to make it one's principle to pursue, in the given circumstances, the end whose source is that incentive.

2.

Kant conceives of an agent's character, her "*Gesinnung*" or "disposition", as an "ultimate subjective ground of the adoption of maxims".¹⁰ An agent's character ties her actions together over time, so that they are not merely a heap of separate and unconnected "willings", and, in Kant's view, this is accomplished by this single basic attitude or commitment that ties together the *maxims* of her actions. He does not, however, think of *Gesinnung* as a natural or cultivated bent in human beings (though at times the text suggests this) but as itself a very general maxim, and so as itself a product of practical reason.¹¹ This creates a regress problem,¹² but Kant's main point is that, for practical purposes, our character is no more an alien cause of our behavior than are natural incentives. For such purposes our *Gesinnung* is not a "given" about which we can only lament or rejoice, and so not feel guilt about nor hope to

change.¹³ But for now it is mainly important to see that an agent's *Gesinnung* is a maxim, not to perform *actions*, but to adopt other more particular *maxims*.¹⁴ As the ultimate ground for the adoption of maxims, we may represent an agent's *Gesinnung* as her ultimate *reason* for acting, a terminus for "Why?" questions regarding her actions.¹⁵

For Kant this terminus is the agent's attitude toward her own happiness and the moral law, more specifically, toward which to privilege in adopting maxims.¹⁶ For instance, my practice of jogging expresses more than a fleeting urge; I think that it will promote my happiness. But how important to me are the demands of happiness in relation to the demands of morality? That, of course, would normally come out only were the two to conflict. But the story would then go in one of two ways: Either I put a premium on happiness or on morality, and that informs my adoption of maxims, even if no conflict actually arises. So, if pressed, and if honest, I will explain and/or justify my practice of jogging by appealing to a deeper interest I have in happiness. If I have a good character, I value morality more than happiness and so put a moral constraint on my pursuit of any end, even one as mundane as fitness. If I have a bad character, then I value happiness more than morality, and so put a prudential constraint on the pursuit of any ends, even those as important as moral ends. As I see it, for this reason Kant's idea of an agent's *Gesinnung* is best thought of, not only as a deep and fundamental explanation of why a person is motivated as

she is, but also as her *deep and fundamental standard of acceptable maxims*.

It might initially strike us as odd to think of a person's character as a "principle", much less as a "standard". Rather, a person's character seems more like something that, at best, is supposed to *explain why* she lives by the principles that she does. Nevertheless, there is an ordinary sense to the idea, as when we think of a person of good character as a person of good principles. And, in the *Anthropology* Kant makes it clear enough that at least at times he has this sort of thing in mind:

Simply to have a character relates to that property of the will by which the subject has tied himself to certain practical principles which [even if "false or defective"] he has unalterably prescribed for himself by his own reason.¹⁷

One's "principles" can be thought of as deep and abiding commitments to an ordering of values in one's life. And, in this sense, a person's character is identified with the principles that express those values. Moreover, we think of both bad and good character in this way. A person of bad character is a person of bad principles.

3.

As I mentioned, I think that there is an ambiguity in the idea of a maxim. The ambiguity is in what sense maxims are supposed to play the role of an agent's "reasons" for acting -- whether they are supposed to figure in the *intentional explanation*, or rather in the *justification and evaluation*, of

actions, or in both.¹⁸ Ultimately, I think, the ambiguity comes from the fact that what Kant *calls* an agent's "maxim" is actually in some cases the former kind of reason, in other cases, the latter. Recognizing this ambiguity is important, for unless we are clear about which of these different aspects of our moral psychology Kant is referring to as a "maxim", we can be left with puzzles such as that concerning IT and weakness. Let me explain, then, why I think Kant uses "maxim" to refer to both "motivating" and "justifying" reasons.

An intentional explanation of an action explains that action, not by way of external physical or psychic pressures that caused her to act as she did, but by way of causes that are "internal" to her, that is, her beliefs and dispositions.¹⁹ Such an explanation thus refers to the agent's motivational system. An agent's reasons in this sense are the beliefs and dispositions that actually motivated her to act on some occasion. And an action that can be explained through her reasons in this way is voluntary and up for moral evaluation. Such evaluations in Kant's view are supposed to focus on the composite of beliefs and dispositions that actually motivated her to act. We do this, in his terms, by construing this composite as a principle or maxim.²⁰ "Maxim" here is Kant's term for what we focus on when we morally evaluate a person's action, namely, what "reasons" actually brought her to perform that action. Notice that we focus here, not on actions in the future or actions that an agent *might* perform, nor necessarily on those

that she thinks *admirable* or *good* to perform, but on actions that have already been, or are being, performed.

Now the beliefs and dispositions that motivate an agent's actions -- those that provide the best construal of her principle or maxim in acting as she does -- may well not reflect what that agent takes to be an adequate justification for acting, even though those beliefs and dispositions were indeed her reasons in the first sense. An agent's reasons in this second sense represent her valuational system. To "adopt a maxim" in this sense is to adopt a standard by which to evaluate and guide one's actions; it is, in short, to *endorse* a way of acting in certain circumstances. Such a principle need not be something so grand as a "moral" or even "prudential" principle. It might be as mundane as "I will make my appearance fashionable and neat in public" or "I will be friendly and charming in social situations". These are principles an agent might endorse and aim to live up to, though she hardly regards them as her moral values, even if they are not in conflict with her moral values. They are values that, to some extent at least, justify her actions from her perspective.

Now, surveying the text, one might think that maxims must always serve the first, motivational role, and so are always the first kind of "reason", namely, principles that actually *bring an agent to act*. This is supported by Kant's claim that a maxim is "a principle on which the subject *acts*",²¹ and his insistence that

when moral value is in question, we are concerned, not with the actions which we see, but with their inner principles, which we cannot see.²²

It is pretty clear that Kant's view is that the moral value of an action that has been performed depends on the motivation, the "inner principle" or "maxim", that actually brought it about. Whether Jane did a morally valuable thing in giving correct change to a customer depends on *what actually motivated her* to do it. By contrast, if Jane merely professes a principle of dealing fairly with customers, but that principle did not actually motivate her, then that principle is not her maxim.²³ This doctrine is based on the normative position that we should not evaluate a person's actions by the values she holds or professes, and in light of which she will feel shame or guilt if she fails to live by them, but by the motivations actually at work in those actions.

If we accept this construal of "maxim", then I could not, for instance, have a maxim to jog to get healthy, say, yet fail to act on that maxim and so fail to jog. For on this interpretation only "objective practical principles", or practical laws, are principles on which an agent may fail to act, being principles on which she *ought* to act. Hence, on this "always motivating" construal of maxims, there would be two classes of principles: Those on which an agent acts, that perhaps sometimes justify but always *motivate* her actions, and that are the focus of moral evaluation; and those on which an agent *really ought* to act, in light of which we should make our

evaluations, and that would *really justify* her actions if she acted on them, but on which she may fail to act.

But if it purports to be exhaustive, this distinction is flawed.²⁴ People may try yet fail to live up to principles that are not objective, not valid for any rational being. But if all maxims must motivate, then such principles cannot be an agent's maxims. Hence, there is no room left for what we ordinarily think of as a person's "principles" -- those standards an agent lives by, her personal codes of conduct, "rules of thumb", and the like, principles that an agent might adopt yet fail here and there to live up to, regardless of whether they are universally valid. Such principles can easily be construed in the *form* of maxims, that is, "I will f whenever in C to get E". They surely conform to the agent's subjective conditions, are valid only for herself or perhaps those who have interests and desires like hers.²⁵ And the agent surely regards these principles as hers, even to the point of identifying with them.²⁶ Moreover, although the text at times does yield the motivational reading, at other times Kant pretty clearly states a different position, for instance, that a maxim is "a principle which the subject himself makes his rule (how he wills to act),"²⁷ and that fully rational beings would always "judge their actions by reference to those maxims" that conform to universal law. These passages seem to indicate something quite different, that Kant regarded maxims as action-guiding principles that an agent adopts for herself, whether or not she lives up to them, and whether or not they really do justify her behavior objectively.

Hence, I think that on Kant's view there are maxims that motivate but don't necessarily justify our actions from our own perspective, and maxims in light of which we justify or condemn our actions but that do not always motivate us. Now, in contrast to this "always motivating" interpretation, I suspect that many Kantians adopt an "always (putatively) justifying" interpretation.²⁸ On this view, whatever I can dignify with the title of being my "principles" must in my view *prima facie* or minimally justify my actions. The considerations I have advanced would thus be better explained by, say, a distinction in levels or generality in principles that are all *prima facie* or minimally justifying from the perspective of agents, but which can come into conflict. For instance, I may have a "deep" guiding maxim of helping others, but when the time comes to help, I see a superficial personal indulgence as justified, and feel guilt for going against a deeper maxim with which I more closely identify. Hence, the principle that I identify with is just a deeper and more general level of justification than the more superficial, but still minimally justifying, maxims on which I sometimes, regretfully, act. Thus, the "always justifying" interpretation would also give us a twofold distinction: Those principles, of varying generality and "depth" that, from an agent's own "subjective" perspective, justify her actions, and those that, from any rational agent's perspective or "objectively", would justify her actions.

But, first, there is a general, and I think insuperable, difficulty for Kant if we relegate maxims to an *always*

justifying role, even minimally. For Kant holds that every action is from some maxim.²⁹ Yet it is clearly *not* the case that every action is justified in the eyes of the agent, even minimally or *prima facie*. Some actions we condemn ourselves for performing, often while we are performing them, and we find nothing at all to be said for them *all the way down*. Otherwise decent people make snide comments to colleagues, emotionally manipulate their own children and blow money they don't have on useless items. And *at least sometimes*, all the while they are doing these things, they condemn their own actions. Nothing at all counts for them in their eyes. But if maxims must always minimally justify actions from the agent's perspective, then we are left with the untenable position that these otherwise decent people really do put some value on these actions after all. Simply put, if every voluntary action is from some maxim, but some voluntary actions are utterly unjustified from the perspective of the agent, then some maxims provide no justification from the perspective of the agent, let alone from an objective perspective.

Second, were we to see such actions as performed from more specific, minimally justifying, maxims that are in conflict with more general "deeper" maxims, we would have to say that, for instance, emotionally manipulating one's children and then feeling guilt afterward involves a *conflict* of values. But this too is a misrepresentation. A conflict in values leads a person to wonder why she values the things that are in conflict. But in this case a person need not wonder this, since from her

perspective she puts no value at all on one of the "conflicting" items. In the interests of charity, then, I think that we should prefer the interpretation I offer, especially given that it fits with ambiguities in the text.

Another interpretation, however, might see these "non-motivating" principles as just *potential* maxims.³⁰ Those principles for or against a given course of life an agent considers, and strives to follow, when she actually acts on them, are her maxims. When she fails to act on them or rejects them, they were simply *potentially* her maxims. But although a distinction of this sort between an agent's potential and actual maxims does make room for some cases, it does not entirely capture the above distinction. For some principles remain an agent's *actual* principles even when she fails to act on them, in the sense that they continue to express her personal values. They are the principles in the light of which she evaluates whether or not she is justified in acting as she does. It is in light of his principle of looking chic that Ed feels as if he has failed when he does not keep his wardrobe current. He thinks that he has failed to do what he should have done *in light of this principle*. Moreover, this is just the sort of principle that agents will often identify with most, as defining "who I am". For this reason, we cannot simply write off an agent as expressing bad faith when she professes principles she fails to live up to. For it is in light of her own principles that we can understand her reactions to her own conduct of guilt, shame or self-satisfaction.

Admittedly, Kant's view, expressed in several places, is that a rational agent cannot will anything other than what she takes to be good in some sense.³¹ And this initially might make it seem that there is in fact no space for a distinction between maxims that *motivate*, as opposed to *justify*, actions. Given this view, it would seem that adopting a maxim *just is* construing oneself as having some justification for acting.

Kant's discussion of acting under the heading of the good in the *Second Critique* reveals a more subtle view, however. Here he explains his differences with what he terms "an old formula of the schools" that we desire nothing except under the heading of the good, and are averse to nothing except under the heading of the bad.³² There is an important condition that must be added to it: It is only "clearly expressed", he says, "when rendered: 'We desire nothing, *under the direction of reason*, except in so far as we hold it to be good or bad.'" ³³ Thus, it is not *impossible* to will something we take to be bad. Rather, it is *impossible rationally* to do so: We are, in other words, irrational to the extent that we will to do something we do not hold to be good. The point here is parallel with regard to Kant's view of our conformity to the Hypothetical Imperative to will the necessary and available means to those ends that we will.³⁴ It is quite possible for us to will an end and yet fail to will the necessary and available means to it; that is why it is so hard to stay on a diet. Our failure to will the means is a failure to will something that we hold to be good (albeit conditionally so, on our willing the end that those means

serve).³⁵ And this failure reveals that in some respect we are practically irrational. We have failed to will in a way that, were our will determined by reason, we would have willed. The same, then, is true of the view Kant identifies as scholastic, namely, it is quite possible for us to will something we hold to be bad, or fail to will something we hold to be good. That is a sign of our irrationality, that our will is not conforming to the "direction of reason".

I think that the above considerations therefore make it reasonable to recognize two conceptions of "subjective" practical principles, along with the single conception of "objective" practical principles or practical laws: There are thus those principles that (a) represent our motivations, and so figure in the intentional explanation of actual actions; (b) represent our valuations, on which we may fail to act and in light of which we will justify (or condemn) our own actions; and (c) are "objective" in the sense that they are valid for any rational agent, and so objectively justify (or condemn) our actions -- principles that any rational agent could accept and endorse. The latter, practical "laws", represent "truths" about practical rationality, and morality in particular. Those principles we adopt as normative represent our system of values, which may or may not match up with what is objectively of value.³⁶ Those that motivate us, however, need not reflect our values at all.

On this view, then, a person can fall short of her own maxims, though voluntarily and intentionally. Hence, I may "make

it my maxim" to jog regularly, but fail to jog; say, instead, I play video games. That I failed does not imply that it is not my maxim to jog. After all, it is in light of that principle that I feel some guilt about my indulgence in video games. But my failure was not accidental, inadvertent or in some other way unintentional. So there is still a maxim that explains my video gaming -- namely, "I will play video games for fun". Thus, "It is P's principle or maxim to ϕ " sometimes refers to P's motivation in ϕ -ing, but at others to P's endorsement of ϕ -ing, regardless of whether she is successful in living up to that endorsement. And this in turn reflects the distinction between an agent's *motivations* and her *values*.

Now the point I want to make here, crucial for what follows, is this: If Kant's idea of a "maxim" refers both to an agent's motivations and her values, then we have a more complex picture of his views concerning the determination of the will, and so concerning IT. In short, IT may be taken as the claim that incentives and the like must be incorporated into maxims in order to be *motivating* reasons, or as the claim that they must be incorporated into maxims in order to be *justifying* reasons. It thus concerns not merely the incorporation of various incentives into our *motives*, but also into our *values*. Moreover, in being our "ultimate" maxim, our *Gesinnung* on this view would represent a final standard of acceptable "values", as well as an internal starting point of motivation.

So Kant's view allows us to will things (irrationally) under the heading of the bad, to act for reasons we take to give

us no justification, and fail to act for reasons we take to give us justification, for acting -- quite apart from whether those reasons really do justify our actions. Indeed, his view allows us to will things we may take to be neither good nor bad, and so to act for a reason that we take to be in some sense indifferent. Yet IT must be at work even in these cases, though not in our *justifications* for acting, since in each case we incorporate an incentive toward the bad or indifferent into a maxim. And if, as I think, Kant has both conceptions of "reasons" in mind under the heading of "maxims", then we must take IT to be at work both in the motivation and justification of behavior.

4.

One reason why it is important to preserve the possibility of weakness is to allow us to distinguish moral weakness from genuine vice.³⁷ Hereafter, I will focus on this kind of weakness, and how it can be explained within Kant's views.

"True vice", Kant says in the *Metaphysics of Morals*, requires that a will "take up what is evil (as something premeditated) into its maxim".³⁸ Thus, as one would expect, Kant's account of vice and virtue rests on IT. More specifically, the genuinely vicious person incorporates into her deepest maxim, her *Gesinnung*, the incentives of happiness and morality in the wrong way:³⁹ In the *Gesinnung* of the genuinely vicious person, the moral incentive is made subordinate to that of happiness, while the reverse is true of the genuinely virtuous.⁴⁰

In the *Religion*, published just before the *Metaphysics of Morals*, Kant distinguished the "empirical" from the "intelligible" character of virtue.⁴¹ His view here is that achieving the latter requires a "revolution" whereby the order of incentives in one's *Gesinnung* is reversed to give priority to the moral law. But the lack of revolution is compatible with "empirical" virtue, or the presence of a steadfast maxim of conforming one's actions to moral law -- the "strength of a man's maxims in fulfilling his duty".⁴² And, by contrast, the occurrence of revolution is compatible with an on-going struggle of reforming one's empirical character, for "there is still a great gap between the maxim and the deed".⁴³ Kant thus acknowledges the possibility that a *good disposition* with the right order of incentives, can still be weak in failing to act accordingly. Moreover, he provides for this possibility as one degree of the human capacity for evil, made actual when an agent adopts "the good (the law) into the maxim of [her] will [*Willkür*], but this good, which objectively...is an irresistible incentive, is subjectively...when the maxim is to be followed, the weaker (in comparison with inclination)." ⁴⁴ This capacity for evil is thus actualized when an agent's establishes the right order of incentives in her will, but fails to be able to put that good will into practice in her particular motivations. And similarly, again in the *Metaphysics of Morals*, Kant says of the lack of "self-governance", that this weakness "in the use of one's understanding coupled with the strength of one's emotions

is only a *lack of virtue*...[and] can indeed coexist with the best will."⁴⁵

Taken together, these passages give us this picture: A *Gesinnung* in which there is a proper order of incentives is quite compatible with a lack of (empirical) virtue. A person's genuinely pure *Gesinnung* may not, in other words, have a sufficient influence on what motivates her in particular situations, though were she fully rational, her own good will would be "irresistible". In effect, she acts against her own best judgment and deepest commitments.

No doubt, much of what Kant has to say about *Gesinnung*, our capacity for evil, and the restoration of the proper order of incentives is entangled with his attempts to understand in moral terms religious doctrines such as providence, redemption, immortality, and original sin.⁴⁶ It may therefore be difficult, initially anyway, to see how these view can be as relevant as I want to portray them to the more mundane problem of weakness. Nevertheless, I think a plausible and common-sense view of weakness can now be pieced together from these discussions. Let us, then, return to the problem with which we started, now focussing on the distinction between weakness and vice.

5.

On the one hand, no incentive can determine an agent's will unless she makes it her maxim to fulfill it; on the other, if an agent makes it her maxim to fulfill an incentive that contravenes a principle she holds dear, then she adopts a new principle that flouts the old. In short, every action is

performed from some maxim, yet some actions go against one's maxim. Thus, there seems to be no room left for someone who exhibits weakness by adopting a maxim and then failing to live up to it. Moreover, IT makes it impossible for someone to act immorally from weakness and not on principle: The most vicious type of coward is someone who performs cowardly actions in some sense *on principle*; she has few scruples about, say, saving herself when she should stand up for others, and suffers few pangs of conscience as a result. ("I'm not risking *my neck!*" she thinks.) Someone who is only weak, however, does feel badly, and does *just because* she acted contrary to her own principles. ("I *should* have stood up for him, but I was so scared!" she thinks.) Kant appears to have no room to see any real difference between the two, since IT implies that the latter person acts on principle after all. At worst, then, there is no weakness at all, just the virtuous and the vicious. At best, however, weakness can only be represented as a kind of erratic behavior - adopting a principle at one moment, only to drop it and adopt another when a more tempting incentive comes along.⁴⁷ No doubt, this too is a failing of practical rationality, but it is quite distinct from the phenomena of weakness. The weak person we are interested in here is not simply a person who *drops and adopts* principles too easily, but is rather one who cannot *live up to* those she adopts and does not drop.

Allison's own view seems to be that, for Kant, the difference between these two cases must just be the presence of self-deception. Moral weakness is a susceptibility "to

subordinate the incentive of morality to that of self-love" which "is self-deceptively taken by the agent as a brute given", thus showing that the morally weak agent's commitment to the moral law "cannot be *fully* genuine".⁴⁸ Thus, the difference between the vicious and the weak is apparently that the weak hide their corrupt principle from themselves. But, though this undoubtedly describes some *purported* cases of weakness, when generalized it implies that the weak are, after all, engaging in principled immorality and so are no different from the vicious. Indeed, they may be worse for not taking responsibility for their choices; weakness is bad faith. This is inadequate to save IT, since it essentially gives up on even trying to respect this phenomena of voluntarily acting against one's own values.

It seems to me that charity should lead us to construe Kant's position, as far as is possible, as respecting this phenomena. And that requires acknowledging that the morally weak are truly committed to morality "down deep". So, first of all, a weak person's *Gesinnung* should be represented as having the right priority established between the moral law and happiness. She really does, in her heart so to speak, put the law first. But she must also voluntarily fail to act according to this "deep down" commitment. So, second of all, wayward incentives determine her choices, and thus she incorporates them into maxims. If my interpretation is correct, then we can now put these elements together to explain this weakness consistently with the truth of IT: There is a conflict, or lack of

integration, between the *values* enshrined in an agent's character and what *motivates* her.⁴⁹

The idea is this. Weak-willed people in general act voluntarily, and so have reasons for what they do -- there is an explanation of their actions by way of their beliefs and dispositions. On Kant's view, we construe these beliefs and dispositions as their "subjective principles" or maxims. However, these maxims are utterly inadequate as judged by their values. And, again on Kant's view, we construe these as "subjective principles" or maxims. By contrast with the weak, the motives of the vicious -- cowards, liars and the like -- are not condemned by their values. Thus, weakness is simply a lack of integration between these two aspects of moral psychology that Kant represents with the single term "maxim".⁵⁰ Her motives cannot be justified by the values she holds dear, and the values she holds dear fail to influence what motivates her.

Moral weakness in the deepest sense results when an agent is motivated in particular cases against the values contained in her *Gesinnung*. In this case, the standard that guides her in making more specific decisions of principle privileges the moral law over the other incentives of happiness. Nevertheless, she voluntarily acts as a wayward incentive inclines her, and so adopts and is motivated by maxims that conflict with her deepest values. For instance, suppose she makes a snide comment to a colleague, against her better judgment. Since her action is voluntary, a principle motivated her to make the comment -- this is required for her action to be explicable in intentional

terms. It does not, however, necessarily call into question the moral integrity of her character. She has not suddenly come to view her interest in making such comments as, after all, a justification for doing so. That would not respect the phenomenon. Instead, she condemns her motivation as valueless, as containing no justification.

A vicious person's *Gesinnung*, by contrast, would put her own personal interests ahead of moral values. That those values forbid such comments, for instance, would be sufficient justification to avoid making them only if that does not conflict with her personal interests. So when the vicious person makes such snide comments, she too is motivated by a maxim of doing so. The difference is that, again "deep down", she does not really condemn that maxim as worthless; her *Gesinnung* is not in conflict with it.⁵¹ She is thus truly contemptuous of others, while the weak person is not, being motivated in a way that has no origin in what she values. And this is just what weakness is: Acting against one's best judgment, for reasons that are insufficient by one's own standards. The weak are thus different from the vicious, not because their motivations are different, but because their *values*, and ultimately, their characters, are different. This is, ultimately, why it is so important that we see IT as addressing not merely how incentives figure in the *motivation* of our actions, but how they figure in our *values* as well.⁵²

Kant claims, however, that this most vicious of persons cannot be *ignorant* of the moral law. That law is part of the

very structure of the will, and so every will must incorporate both it and happiness into its *Gesinnung*.⁵³ Nevertheless, his view only *requires* that the vicious person recognize its authority by condemning her deepest values -- condemning the priorities enshrined in her *Gesinnung*. In so doing, she does not yet condemn her *motives*. This sets her apart from the weak, who condemn their motives in light of their *values*. The vicious person must first change those values before she can condemn her motives.⁵⁴

6.

IT, then, does not make weakness impossible. Even though the weak and the vicious both incorporate wayward incentives into their motives, this does not destroy nor trivialize the difference between them. The difference is in how they incorporate incentives into their values, not their motivations.

NOTES

1. Thanks to Marcia Baron, Nelson Potter, Andrews Reath, Alexander von Schönborn, Mark Timmons and a helpful referee for comments and discussions on these topics.

2. R 19. Except for *Religion* and the first *Critique*, paginations refer to the volume and page numbers from the Prussian Academy edition. References to the first *Critique* are to the pages of the A and B editions, *Religion* to the translator's pages. The following are the translations I use:

- A *Anthropology from a Pragmatic Point of View*. V. L. Dowdell, tr. H.H. Rudnick, rev. Carbondale & Edwardsville, IL: S. Illinois U. P., 1978.
- CJ *Critique of Judgment*. J. C. Meredith, tr. Oxford: Oxford U. P., 1952
- G *Groundwork of the Metaphysics of Morals* H. J. Paton, tr., New York: Harper, 1964
- CPR *Critique of Pure Reason*. N. Kemp Smith, tr., New York: St. Martin's Press, 1965.
- CPrR *Critique of Practical Reason*, L. W. Beck, tr. Indianapolis: Bobbs-Merrill, 1956.
- MM *The Metaphysics of Morals*. M. Gregor, tr. New York: Cambridge University Press, 1991

R *Religion within the Limits of Reason Alone*. T.H. Green and H. H. Hudson's trs. New York: Harper & Row, 1960.

³. In his *Kant's Theory of Freedom*. (Cambridge: Cambridge University Press, 1990); e.g., p. 40.

⁴. Marcia Baron first raised a version of this problem in a symposium on Allison's views in "Freedom, Frailty and Impurity." *Inquiry* 36 (1993): 431-4.

⁵. Though she cannot, of course, incorporate evil *for its own sake* into her maxim. (See, e.g., R 31). But Kant is fairly explicit about this difference between "true vice", in which a person acts contrary to duty on principle, and mere "frailty" or weakness, in which a person merely "lacks virtue" and so acts contrary to, or fails to act on, her genuinely avowed moral principles. (MM 6:408; R 24-5) Although frailty represents a character defect, no doubt, Kant reserves the term "vice" for principled immorality.

⁶. See, e.g., CPR A532/B561 -- A559/B587; G 4:446-454; CPrR 5:29-31, 51-8, 72-90.

⁷. See, e.g., CPR A536/B564 -- A 542/B570; G 4:448; CPrR 5:3-16. However, note two things. First, "desire" is an equivocal term; even Kant can agree that every voluntary action is motivated by desire in the trivial sense that a voluntary action is one that the agent performed because he desired to. But Kant would *deny* that every voluntary action is motivated by "desire" in the

sense of an empirical psychological state. See G. F. Scheuler. *Desire* Cambridge, MA: MIT Press, 1995; e.g., pp. 45-52. Second, whether it is a *desire* that causally determine us from a theoretical perspective would depend on further argument identifying causes as desires.

⁸. See, e.g., Beck, L. W. *A Commentary on Kant's "Critique of Practical Reason."* Chicago: University of Chicago Press, 1960, pp. Bittner, Rüdiger. "Maximen" in *Akten des Kongresses*, G. Funke and J. Kopper, eds. Berlin: de Gruyter, 1974, pp. 485-98; Höffe, Otfried. "Kants kategorischer Imperativ als Kriterium des Sittlichen," in *Ethik und Politik*, Höffe, ed. Frankfurt: Suhrkamp, 1979, pp. 84-119; O'Neill, Onora. "Universalization and Consistency" in her *Constructions of Reason* (NY: Cambridge U. P., 1989); Allison, op. cit. 1990, pp. 85-94.

⁹. See G 4:400n, 421n.

¹⁰. See, e.g., R 20; for some recent discussion of Kant views on character, see, e.g., op. cit. Allison, pp. 129-79; Ameriks, Karl. "Kant on the Good will." in *"Grundlegung zur Metaphysik der Sitten": Ein Kooperativer Kommentar*, O. Höffe, ed, Frankfurt: Vittorio Kolstermann, 1989, pp. 45-65; Herman, Barbara, *The Practice of Moral Judgment*. Cambridge, MA: Harvard U. P., 1993, esp., pp. 1-44, 208-40; Baron, Marcia. *Kantian Ethics Almost without Apology*. Ithaca: Cornell U. P., esp. 205-226; O'Connor, Daniel. "Good and Evil Disposition." *Kant Studien* 76 (1985), pp. 288-302.

¹¹. See, e.g., R 20-1

¹². R 20-1. Since if every choice is based on a maxim, and we choose our disposition, then we must have some further maxim for that choice, and so on.

¹³. For more discussion of this point, see my "Kant's Conception of Virtue", *Jahrbuch für Recht und Ethik/Annual Review of Law and Ethics* Vol. 5, forthcoming, 1997.

¹⁴. See my *ibid.*

¹⁵. R. 20

¹⁶. R. 25, 31

¹⁷. A 8:292

¹⁸. For a discussion of this distinction, see, e.g. Raz, Joseph. ed., *Practical Reasoning* Oxford: Oxford U. P., 1978: pp. 2-4. Parfit, Derek. *Reasons and Persons*. Oxford: Oxford U. P., 1984; p. 118.

¹⁹. Note that I use "disposition" here, not in the sense of *Gesinnung*, but in the sense of "motive". I say "beliefs and dispositions", rather than "beliefs and desires", because this leaves open the idea that an intentional explanation in terms of reason might appeal to *rational dispositions* rather than simply "beliefs" alone. And I make no assumptions here about *what kind* of causality is required. And, again, the sense of "desire" that plays a role in intentional explanations seems to be more philosophical than the ordinary sense of that term. It seems to mean just "disposition". On this, see, e.g., Nagel, Thomas. *The Possibility of Altruism*. Princeton, NJ: Princeton University Press, 1970, pp. 27-32; and *op. cit.* Scheuler, pp. 43-77.

20. I take it that the *beliefs* of the agent would be that she is in circumstances C, and that her ϕ -ing will bring about or realize some end E; her *desires* would then be represented by their content, or E. From these, we then get the basic maxim formula of "I will ϕ in C in order to bring about or realize E". For a recent examples of this way of reading Kant, see, e.g., Thomas E. Hill, Jr., "The Rationality of Moral Conduct", in his *Dignity and Practical Reason*. (Ithaca: Cornell U. P., 1992): 97-122, p. 106; Latham, Noa, "Causally Irrelevant Reasons and Action Solely from the Motive of Duty", *The Journal of Philosophy* (1994): 599-618.

21. G 4:421n.

22. G 4:407

23. We do not entirely ignore this, of course. After all, an agent might have been *trying* to act on a good maxim, and that may count some when she fails.

24. See Allison's discussion of other controversies, in *op. cit.*, 1990, 85-94.

25. See, e.g., G 4:421n

26. See CPrR 5:19

27. "... (wie es nämlich handeln will)" MM 6:225; thanks to an anonymous referee for this citation.

28. Thanks to an anonymous referee for pressing this objection. See, e.g., Rawls, John "Themes in Kant's Moral Philosophy", in *Kant's Transcendental Deductions*. E. Förster, ed. (Stanford, CA: Stanford U. Press, 1989); pp. 82-3, and 254 n.2.

²⁹. Aside from IT itself, see, e.g., G 4:412

³⁰. See, e.g., Nell, Onora. *Acting on Principle*. NY: Columbia U. Press, 1972; p. 34 n.7.

³¹. See G: 4:414-417; CPrR 5:58-68, CJ 5:209. Cf. Stocker, Michael, "Desiring the Bad." *Journal of Philosophy* " 76; pp. 738-53.

³². CPrR 5:58-61; I take it here that, in this discussion, the "faculty of desire" Kant refers to is simply the will.

³³. CPrR: 5: 60; my emphasis.

³⁴. G 4:414-15; see Thomas E. Hill, jr. "The Hypothetical Imperative" *Philosophical Review* 82 (1973): 429-50.

³⁵. G 4:414: "Every practical law represents a possible action as good...If the action would be good solely as a means to *something else*, the imperative is *hypothetical*."

³⁶. See also, e.g., op. cit. Nell, and op. cit. O'Neill, 1989.

³⁷. This is the case of weakness that most interests Baron. See op. cit.

³⁸. MM 6:408

³⁹. R. 31-2

⁴⁰. R 31

⁴¹. R 12-3, 42-3

⁴². MM 6:394

⁴³. R 42

⁴⁴. R. 25

⁴⁵. MM 6:408; "...doch ist diese Schwäche im Gebrauch seines Verstandes, verbunden mit der Stärke der Gemütsbewegung, nur

eine Untugend und gleichsam etwas Kindisches und Schwaches, was mit dem besten Willen gar wohl zusammen bestehen kann..."; see also MM 6:384.

⁴⁶. See, for instance, Allen Wood's illuminating discussion in *Kant's Moral Religion*. Ithaca: Cornell U. P., 1970; pp. 208-248.

⁴⁷. At best, then, it would recall Plato's description of the "democratic personality" in *Republic* VIII, 558c - 562a.

⁴⁸. "Kant on Freedom: A Reply to my Critics" in *Idealism and Freedom*, Cambridge, 1996; p. 121.

⁴⁹. See Phillip Pettit and Michael Smith, "Practical Unreason", *Mind* (1993): 53-79, for a rich discussion of the varieties of lack of integration that can occur between an agent's values and motives.

⁵⁰. Mark Timmons (in "Evil and Imputation in Kant's Ethics", *Jahrbuch für Recht und Ethik*. Band 2 (1994) pp. 125-7) seems also to see IT as applying separately to justifying and motivating reasons. But the terminus of both, for Timmons, is in the agent's *Gesinnung*. So the full rationalizing explanation for moral weakness implies a "violation of the principle of autonomy at the level of one's disposition." But then Timmon's solution fails as well. For if a weak person's *Gesinnung* is evil, then she must have inappropriately incorporated incentive into her *justifying* reasons as well, since a person's *Gesinnung* is her *ground* for adopting maxims. So on his view, the weak are, after all, principled in their immorality; they have not acted against their own deepest commitments, but have put them into practice.

⁵¹. Notice that the genuinely vicious person may well act impeccably. For she may see this as the best way to satisfy the demands of desire.

⁵². I have argued elsewhere that this allows Kantians an alternative explanation of phenomena that Aristotelians make use of -- the view that our dispositions "color" our moral perception. For a Kantian, the problem is not perceptual; rather, the problem is in the standard by which an agent evaluates her reasons. See my op. cit. "Kant's Conception of Virtue".

⁵³. See, e.g., R 31

⁵⁴. This is, in turn, why a change of character seems mysterious in Kant's hands: What justification could a bad character have to reform herself, if her character determines what she takes to be justification for acting in the first place? This suggests an objection. For if the vicious person must condemn her disposition, then she cannot fail to condemn her maxims as well. Why, in other words, should a person who really does see herself as having a bad character think that she nevertheless has justification to act in the evil ways that she acts? Put another way, how *else* can a person come to regard her character as bad *except* by recognizing the fact that she acts with no justification?

Here is the difference: What the person of bad character focusses on, if she is reflective, is not the fact that she is guided by inadequate maxims. It is the deeper issue of her

priorities that is in question. A person of weak character, however, will be focussed, not on her priorities (though if she is chronically weak, she may come to question these as well), but on her ability to put those priorities into practice. It is also a difference that will be reflected in practical terms, concerning how each is to improve: A person with bad character must come to see that she must change the way that she evaluates her options. For the weak, it is not how they evaluate their options that is in question, but, being confident of their values, they cannot seem to bring themselves to act accordingly..