



3-2002

Kant's Taxonomy of the Emotions

Kelly Sorensen

Ursinus College, ksorensen@ursinus.edu

Follow this and additional works at: http://digitalcommons.ursinus.edu/philrel_fac

 Part of the [Ethics and Political Philosophy Commons](#)

Recommended Citation

Kelly D. Sorensen (2002). Kant's Taxonomy of the Emotions. *Kantian Review*, 6, pp 109-128. doi:10.1017/S136941540000162X.

This Article is brought to you for free and open access by the Philosophy and Religious Studies Department at Digital Commons @ Ursinus College. It has been accepted for inclusion in Philosophy and Religious Studies Faculty Publications by an authorized administrator of Digital Commons @ Ursinus College. For more information, please contact aprock@ursinus.edu.

Kant's Taxonomy of the Emotions **Kelly D. Sorensen**

If there is to be any progress in the debate about what sort of positive moral status Kant can give the emotions, we need a taxonomy of the terms Kant uses for these concepts. It used to be thought that Kant had little room for emotions in his ethics. In the past three decades, Marcia Baron, Paul Guyer, Barbara Herman, Nancy Sherman, Allen Wood, and others have argued otherwise.¹ Contrary to what a cursory reading of the *Groundwork* may indicate, Kant thinks the emotions play an important role in the moral life. I want to extend the work of Baron, Guyer, Herman, Sherman, and Wood in three ways. First, I will diagram Kant's taxonomy of feelings and emotions. Agreement on such a taxonomy should make it easier to evaluate debates about Kant and the emotions. Second, I will focus on a certain subclass of emotions – reason-caused affects – that have previously received little attention, even by these Kant scholars. Third, these scholars base much of their defense of Kant on his later works – especially the *Metaphysics of Morals* (1797) and the *Anthropology* (1798); but Kant's fairly rich taxonomy of the emotions, including reason-caused affects, is clearly in place at least as early as the *Critique of Judgment* (1790).² I believe that the *Critique of Judgment* is an importantly ignored resource for understanding the moral role of the emotions for Kant. The third *Critique* makes positive, philosophically interesting claims about the emotions and morality. Kant emphasizes certain roles for emotions in this work that he develops to the same extent nowhere else. Nevertheless, the *Critique of Judgment* goes all but unmentioned by many who write on these issues. In what follows, I will defend as many of my claims as possible using the third *Critique*.

I Kant's Taxonomy of the Emotions

To say that Kant has a "theory" about the emotions would be a misstatement for at least two reasons. The first reason is terminological. As I will use the term in this paper, "emotions" is a *general* term for all the states designated by words such as affects, feelings, some inclinations and desires, passions, etc. Kant uses no single German word to play this role. It is true that he uses terms that are often translated as the English word "emotion:" *Gefühl*, *Affekt*, and *Rührung*, for instance. But for Kant none of these words plays the same role – a shorthand term for the whole family of related concepts – that "emotion" will play in this paper. Take *Rührung* as an example. Often rendered as "emotion" in English translations of Kant,³ *Rührung* is only a subclass of what I am calling "emotions." *Rührung* requires a sort of mental movement – a shift or alternation between, say, an excess and deficiency of some internal force (K3 5:226). Other less complex or more static feelings, such as simple liking or attraction, are not *Rührungen*. By using a general term like "emotion" that Kant does not, I will avoid pre-empting any of his own terms.

A second reason it is wrong to say that Kant has a "theory" of emotions is that he never gives a general, overarching account of them in a single place. This is not to say that he is haphazard when he writes about emotions. On the contrary: Kant seems to be working from an impressively consistent taxonomy of desires, feelings, and the relations between them. He also has long lists of affects and passions and the distinctions among them. There is considerable evidence that Kant intends the words he uses for particular states or faculties related to emotions – inclinations (*Neigungen*), affects (*Affekten*), passions (*Leidenschaften*), desires (*Begierden*), etc. – as *technical* terms. Rich, specific definitions and examples for these terms occur in the

Anthropology, and Kant seems to use them *consistently* throughout his critical and post-critical career. I will not be claiming that all of this adds up to a full "theory" of emotions in the moral life. I would only claim that he has views about the emotions that are important to understanding his ethics – views that are often misunderstood. I want to begin by outlining these views.

It is important to understand why Kant is thought to limit the role of emotion at all. Many misunderstandings about Kant and the emotions arise from his use of the term "inclination" (*Neigung*). "Inclination" is Kant's technical term for "habitual sensible desire (*habituelle sinnliche Begierde*)" (A 7:251).⁴ Human beings find themselves with all sorts of empirical or sensible desires – "empirical" in that these desires exist independently of any reasoning or reflection. Inclinations include desires for food and rest, the desire for self-preservation, and love and sympathy for others. Inclinations cover a broad spectrum of human desires, and clearly only some of these (such as love and sympathy) fall under the category of "emotions" (as I am using the term). Kant's opponents sometimes argue that since Kant assigns no special moral esteem to inclinations, he has no room for emotions such as love and sympathy. The important mistake in this accusation will soon become clear.

It is in fact the case that Kant assigns no special moral esteem to inclinations. Beginning even before the *Groundwork*, Kant consistently gives two reasons that "inclinations" are not a suitable foundation for morality. First, inclination cannot be a reliable *criterion* or "measuring rod" for morality, since one may be inclined to do what is not right. Second, inclination cannot be a reliable *motivation* or "mainspring" for morality, since there will certainly be occasions where one may not be inclined to do what is right. In both cases, note that it is inclination's *unreliability* that rules it out as a candidate ground for morality. Kant distinguishes these two issues as early as the *Lectures in Ethics* in 1780: "We must distinguish between measuring-rod

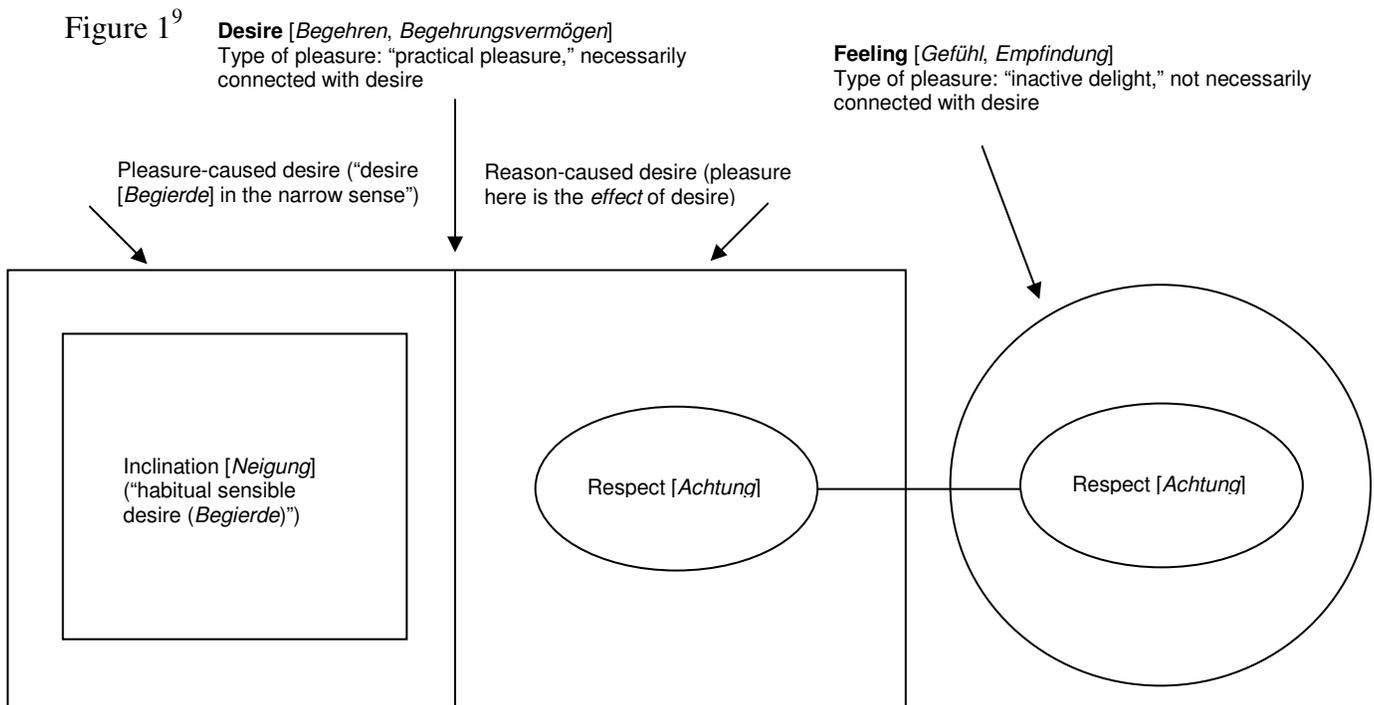
and mainspring. The measuring-rod is the principle of discrimination, the mainspring is the principle of the performance of our obligation."⁵

There is a third and fourth crucial reason that inclination has a questionable moral status. These reasons are perhaps less obvious in the *Groundwork*,⁶ but far more important than the first two in Kant's other writings. The third reason is this. For Kant, inclinations are not the products of an active, free will, but rather the products of deterministic nature that we possess passively without any activity of our own. As such, they are not suitable objects for moral esteem. Why attribute special esteem to inclinations when their possession is a mere matter of luck? Key for Kant here is the notion of *freedom*: only action according to certain principles that one's free will actively formulates and adopts may have "moral worth" (Kant's term for special esteem). Actions have moral worth only insofar as they arise from the active agency of human agents, not as passive side effects of nature.⁷ The fourth reason concerns the relation willing has to the moral law. One can take inclination as a reason for action in the absence of any antecedent commitment to morality; but actions have special esteem only insofar as they are connected to some antecedent commitment to morality. As Kant says in the third *Critique*, a characteristic of human morality is that "reason must exercise dominion over sensibility" (K3 5:269). The third and fourth reasons are in fact his deepest reasons for denying inclination a positive moral status.

These are four strong reasons against assigning inclination either a determining role or a special positive status in morality. But to deny *inclination* these moral roles is not to deny *all* emotions a moral role. To see why, we will need to understand Kant's taxonomy of these and other terms. It is to that taxonomy that I turn now.⁸

Kant believes the mind consists of three general faculties or powers. He makes this most clear in the table of faculties included in the Introduction to the third *Critique*. The three

faculties are *cognition*, *feeling* (of either pleasure or displeasure), and *desire* (K3 5:198). The last two faculties – and crucially, the relation between them – are important for understanding Kant's views on emotions. The following diagram will help make this relation more clear. In the diagram, desires are represented by rectangles, feelings by ovals and circles. As I will explain below, since certain feelings are necessarily connected with some desires, and since Kant refers to these specific desires by the name of the feeling associated with them, it will be helpful to represent these feelings in a second place: among the desires. In such a case, a solid line connects the two representations.



"Feeling" (*Gefühl, Empfindung*) is the capacity to be susceptible to pleasure and pain (MM 6:211). "Feeling" is also Kant's term for specific *instances* of this susceptibility to pleasure and pain – that is, for specific pleasures and pains, not just to the capacity to have them.¹⁰

Susceptibility to sensible feeling is absolutely necessary to any human experience. This point is implicit in the second *Critique* and the *Metaphysics of Morals*. There, Kant points out that certain specifically moral feelings like "respect" depend on a deeper susceptibility to pleasure and pain: "*Sensible feeling*, which underlies all our inclinations, is indeed the *condition* of that feeling we call respect" (K2 5:75; emphasis mine). But the strongest statement of the necessity of this sensible feeling is in the third *Critique*. Kant says that specific feelings of pleasure have become so mixed with our cognition that we have forgotten they exist; but the pleasure of, for example, "being able to grasp nature and the unity in its division into genera and species" is what "alone makes possible the empirical concepts by means of which we cognize nature in terms of its particular laws." "*Even the commonest experience would be impossible without it* [that is, this pleasure]" (K3, 5:187).¹¹ Susceptibility to sensible pleasure and pain, then, is a condition at the deepest root of human experience.

The faculty of "desire" is distinct from feeling, but related to it in an important way. In fact, the relation between feelings and desires allows us to distinguish two types of feeling: (1) those feelings *not necessarily* connected with desire and (2) those feelings *necessarily* connected with desire (see Figure 1).

Consider the first sort of feeling. For feelings not necessarily associated with desires, pleasure "attache[s] only to the representation by itself." Kant calls this pleasure "inactive delight," and the feeling associated with it "taste" (MM 6:212). Feelings of this sort are characteristic of the experience of the beautiful described in the third *Critique*. To experience something as beautiful is to represent it to oneself *disinterestedly* – that is, to find one's representation of an object agreeable without necessarily taking an interest in the existence of the

object (K3 5:222). One can experience this sort of pleasure without trying to make some change in the physical world.

Now consider the second sort of feeling – the sort necessarily connected with desires. In this case, the pleasure attaches "to the existence of the object," not (or not *just*) to one's representation of it (MM 6:211). Kant calls this sort of pleasure "practical pleasure," since the desire associated with it aims at the existence of some object or state of affairs (MM 6:211). This is what Kant means when he describes the faculty of desire (*Begehren*, or *Begerungsvermögen*) as "a being's faculty to be by means of its representations the cause of the reality of the objects of these representations" (K2 5:10n, emphasis removed; MM 6:211, A 7:251). To desire something, then, is to represent it with a feeling of pleasure and to seek to bring it about. So far this framework of desires and feelings is only schematic, but one can already see that at least *basic* desires and feelings play an inescapable role in Kant's psychology of moral action: *no* moral willing or desiring will be possible without feeling of some kind. As Kant says in the third *Critique*, "the attainment of an aim" – moral or otherwise – "is always connected with the feeling of pleasure" (K3 5:187).

Desire necessarily involves feelings, but desire is of two sorts, depending on whether the pleasure associated with it is the *cause* of the desire or instead its *effect* (K3 5:221-222; MM 6:212). Kant calls *pleasure-caused desire* "desire [*Begierde*] in the narrow sense" (MM 6:212). In this case, an agent seeks to bring about the existence of some object or state of affairs because of some antecedent pleasure. When these desires are habitual, Kant gives them their own term: inclinations.

In the case of the other sort of desire, pleasure is the *effect* of the desire. Here it is *reason* that causes the desire, which in turn results in pleasure. As early as the *Groundwork*, Kant

recognizes the existence of at least one such *reason-caused desire*: he calls the feeling necessarily connected with it "respect" or "moral feeling."¹² "Though respect is a feeling," he says, "it is not one *received* by means of influence; it is, instead, a feeling *self-wrought* by means of a rational concept" (G 4:401n; see also K2 5:72-80). Kant's treatment of respect and moral feeling in the third *Critique* is consistent with the characterization of these emotions in the *Groundwork* and the second *Critique*. For instance, the third *Critique* indicates that "in the case where an underlying a priori principle in reason determines the will," the associated "pleasure (in moral feeling) is the consequence of that principle" (K3, 5:289). We have a "an ability for determining a priori with regard to mere forms of practical maxims (insofar as such maxims qualify themselves for giving universal law) a liking that we make a law for everyone; this judgment is not based on any interest, *yet it gives rise to one*" – namely, "moral feeling" (K3 5:300; the italics are Kant's). So the third *Critique*'s treatment of reason-caused desire is consistent with Kant's treatment of it elsewhere.

The class of "pleasure-caused desires" – desires where an agent seeks to bring about some object or state of affairs based on some pre-existing pleasure – has a subclass for those pleasure-caused desires that are *habitual*. Kant's technical definition of inclination captures this: inclination (*Neigung*) is "habitual sensible desire" (A 7:251; MM 6:212). As we saw above, inclinations are not only unreliable as a motivation ("mainspring") and standard ("measuring rod") for morality, but more importantly, actions based on them are not necessarily the result of an agent's free willing in connection with the moral law. The *habituality* of the desires that are inclinations seems to be especially worrisome to Kant: when an agent habitually acts on the same pleasure-caused desire, we may say of her that she has a sort of rule or principle of action (a "maxim" – G 4:399-400) by which she acts. Kant elsewhere describes an inclination as "a

sensible desire (*sinnliche Begierde*) that serves the subject as a rule (habit)" (A 7:265).¹³ This rule or habit may be neither actively chosen nor connected to a commitment to morality, and this is precisely why acts based on it are not worthy of special esteem or moral worth, even if these acts are in accordance with the moral law: agents with inclination-based maxims are simply in the habit of following antecedent pleasures. By contrast, agents whose maxims are based on active, freely willed respect for the moral law are worthy of esteem.¹⁴

The third *Critique* is the first place where Kant makes two new claims about the role of emotions in morality. We saw above that susceptibility to feelings of pleasure and pain was a condition of *experience*. But Kant makes a stronger, more specific claim in the third *Critique*: susceptibility or "the predisposition to the feeling for (practical) ideas, i.e. to moral feeling" is a condition of *morality*, not just a condition of experience (K3 5:265). Just as there is no experience without the capacity for pleasure and pain, there is no morality without the capacity for the specific feeling of respect or moral feeling. As Kant will say later, it is a misunderstanding to think anyone could have a *duty* to acquire these sort of feelings, since they are "*subjective* conditions of receptiveness to the concept of duty" (MM 6:399).

The second new claim about the emotions in the third *Critique* is this: "respect" or "moral feeling" is not the only sort of emotion that reason can produce. According to the *Metaphysics of Morals*, reason can also produce the following four feelings: conscience, love of human beings, moral feeling, and respect (MM 6:399-403). The latter two are understood by 1797 as distinct from one another: "moral feeling" is a feeling of approval toward actions required by duty or a feeling of disapproval toward actions prohibited by duty; "respect" is a combination of *both* pleasure and pain directed toward either a person or the law itself. But the third *Critique* contains evidence that Kant thought reason could produce a variety of emotions seven years

before the *Metaphysics of Morals* was published. These new reason-produced emotions include some of the emotions Kant calls *affects*. In order to understand them, we will need to enrich the diagram about desires and feelings.

II Reason-Caused Affects

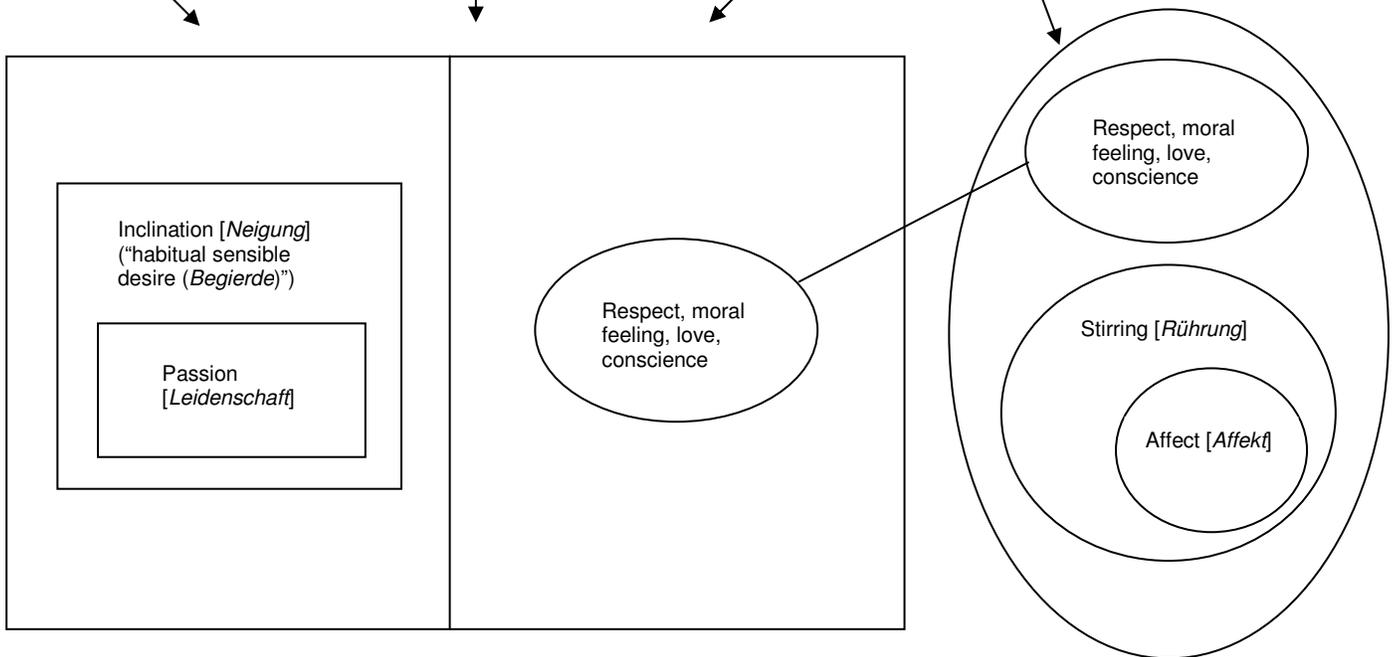
Figure 2

Desire [*Begehren, Begehrungsvermögen*]
 Type of pleasure: "practical pleasure," necessarily connected with desire

Feeling [*Gefühl, Empfindung*]
 Type of pleasure: "inactive delight," not necessarily connected with desire

Pleasure-caused desire ("desire [*Begehrde*] in the narrow sense")

Reason-caused desire (pleasure here is the *effect* of desire)



Several new subclasses of desires and feelings appear with their associated technical names in Figure 2. Within inclinations is another subclass of desires much more dangerous to morality: *passions* (*Leidenschaften*). Passions are *reflectively integrated* habitual sensible

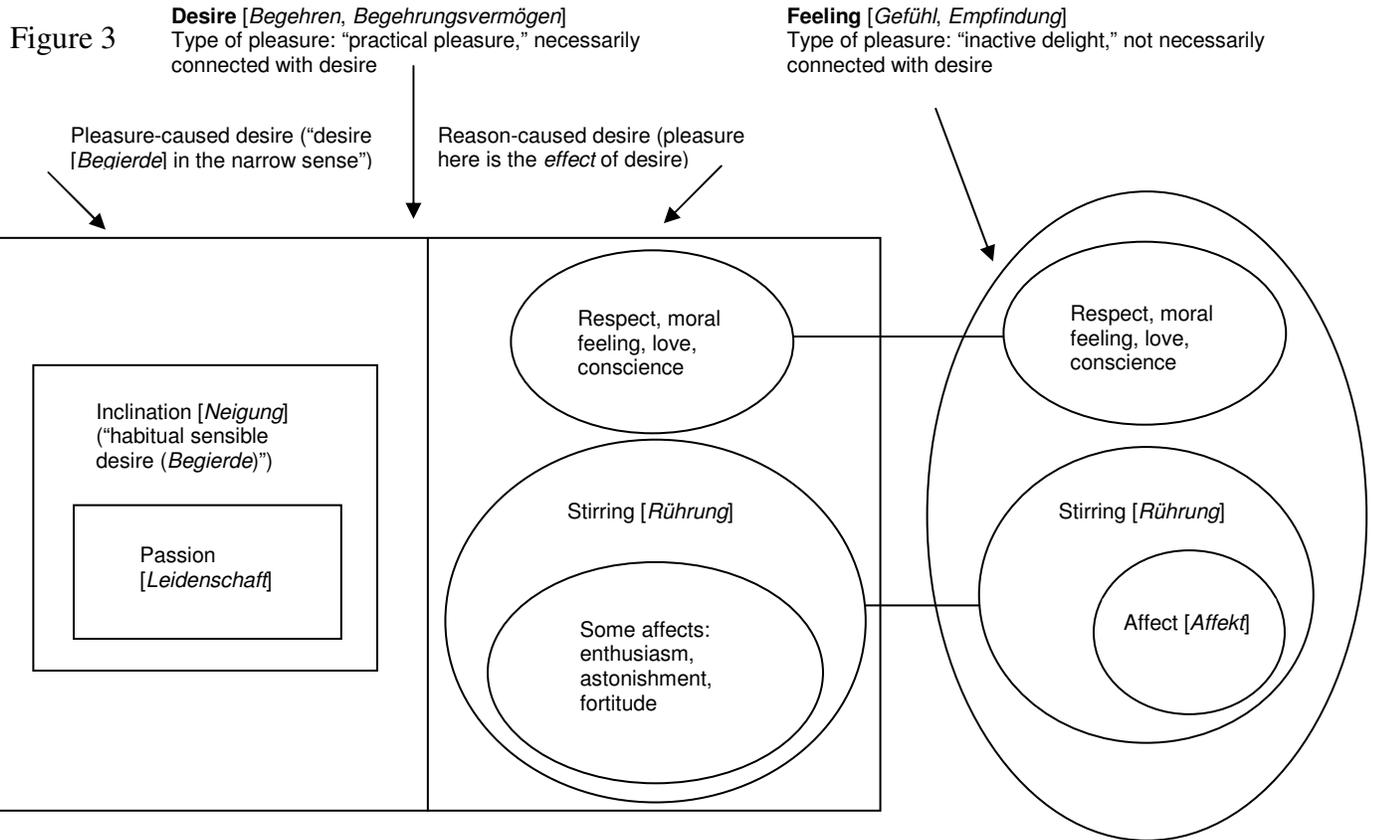
desires. Some of Kant's comments suggest that passions are simply very strong, or very long-lasting, inclinations (A 7:251, MM 6:408). But more characteristic of passions is the agent's *reflectiveness* and the broad *integration* of the passions into the agent's maxims. Passions are not only persistent and habitual (like inclinations), but also *deliberate* (K3 5:272n). An agent who calmly and deliberately acts on a passion reflectively lets inclinations constitute his maxims. "The calm with which one gives oneself up to it [i.e., a passion] permits reflection and allows the mind to form principles upon it and so, if inclination lights upon something contrary to the law, to brood upon it, to get it rooted deeply, and so to take up what is evil (as something premeditated) into its maxim" (MM 6:408). The metaphors Kant uses to describe passions stress the pervasiveness and integrality with which they attach to an agent's maxims: passions are "like a stream that burrows ever deeper in its bed" (A 7:252), and like chains that have "already grown together with his limbs, so to speak" (A 7:267). As ever, Kant's principal worry is the effect of this class of desires on freedom: once passions are firmly in place, "the mind's freedom . . . is abolished" (K3, 5:272n). Other desires and feelings may need to be under the control of reason; but passions "are, without exception, evil as well" (A 7:267).

We have been examining a subclass of "desires." But Kant also believes there are subclasses of *feelings* that are not necessarily connected to desires. One such broad subclass – a subclass Kant emphasizes in the third *Critique* – includes *Rührungen* (which I will translate as "stirrings").¹⁵ Stirrings are constituted by a movement from or alternation between one feeling and its opposite. This makes them more complex than, for example, aesthetic feelings, which require no such movement between opposites (K3, 5:222, 5:245, 5:258). Kant describes "stirrings" in the third *Critique* as "sensation[s] where agreeableness is brought about only by means of a momentary inhibition of the vital force followed by a stronger outpouring of it" (K3

5:226). A parallel definition follows later in the same work: stirrings include "a pleasure that arises only indirectly: it is produced by the feeling of a momentary inhibition of the vital forces followed immediately by an outpouring of them that is all the stronger" (K3 5:245). Although Kant talks about a "stronger" outpouring of an originally inhibited force, not all stirrings are powerful or violent. "We have *courageous* stirrings [*mutig Rührungen*], and we have *tender* ones" (K3 5:273).

If stirrings reach a certain level of strength, they can become *affects* [*Affekten*].¹⁶ The "stirrings [*Rührungen*] that can reach the strength of an affect are very diverse" (K3 5:273).¹⁷ Stirrings become affects when they *impede reflection*. An affect is "a feeling of pleasure or displeasure in [a subject's] present state that does not let him rise to *reflection* (to rational consideration of whether he should give himself up to it or refuse it)" (A 7:251). The strength of an affect plays a role in its impedance of reflection, but "generally speaking, what constitutes a state of affect is not the intensity of a certain feeling but rather the lack of reflection that would compare this feeling with the totality of all the feelings (of pleasure and displeasure) that go with our state" (A 7:253-254). Affects are sudden and overwhelming, "like water breaking through a dam" (A 7:252), but they never last long (MM 6:408, K3 5:272n). Kant helpfully and consistently contrasts affects with passions in the three major texts where he discusses them. Affects prevent reflection, but passions are insidiously compatible with it; affects are sudden, short and "open"; passions "cunning," "hidden," and long lasting (K3 5:272-273, MM 6:408, A 7:251-282).

Let me make one final change to the diagram:



New in Figure 3 are stirrings and affects *within* the class of reason-caused desires. Both passions and affects can be threats to the sovereignty of reason, but while passions always entail evil and vice (A 7:251, 7:267), affects "can indeed coexist with the best will" (MM 6:408). This last point suggests that Kant (at least by 1797 when he published the *Metaphysics of Morals*) regarded affects less negatively than passions. In fact, Kant sees a *positive* role for affects in the *Anthropology*, published just a year later. Reason can produce not only a handful of feelings such as moral feeling and respect (MM 6:399), but a variety of affects as well.

Two points about the notion of reason-caused affects are in order. First, "moral feeling" or "respect" seems to be the only sort of emotion that we can know about *a priori* (K2, 5:73). On

Kant's model of moral psychology, reason must produce action by producing desires in agents; we can know this independent of experience. Affects, by contrast, are the sort of emotion we can only know about *a posteriori*. With this in mind, it is no surprise that Kant would not talk about *a posteriori* emotions in the *Groundwork* and the second *Critique*. His purpose in those works is to discuss the *a priori* foundations of morality, not the possible *a posteriori* effects of reason. Second, it is an open question whether Kant thinks reason can produce affects *ex nihilo*, or whether it simply stirs up affects that are already latent within us. Kant does say that reason can "cause" an affect (A 7:254), suggesting the former. But there is more evidence that he thinks the latter. Kant says that reason "arouses" (A 7:257, 7:269) and "stirs up" (A 7:261) affects; and he advocates the "cultivation" of feelings like sympathy that "nature has implanted in us," where doing so may require us to "expose ourselves" to the subject matter of the "*humanity*" or humanities (K3 5:355), or to seek out "the poor who lack the most basic necessities" (MM 6:457). Also unclear is the *extent* to which Kant thinks reason can control the affects it arouses.

I wish to discuss three of the most interesting reason-caused affects. The first is a natural outgrowth of moral reasoning; the second two have stronger and unambiguously positive roles in morality. I will examine what Kant says about these affects in the later works (the *Anthropology* and the *Metaphysics of Morals*); then I will look for evidence of them in the third *Critique*. As it turns out, the first two *do* clearly appear in the third *Critique*; and Kant had the taxonomical resources to include the third, had he wished.

The first affect is *enthusiasm* [Enthusiasmus] – a sort of energetic excitement about morality. "When reason represents the morally good it can enliven our volition," creating an "*enthusiasm* [Enthusiasmus] of good intentions" (A 7:254). Although two of Kant's terms – *Enthusiasmus* and *Schwärmerei* – are often both rendered as "enthusiasm" in English

translations, it is crucial to distinguish them.¹⁸ Kant seems to think that *Enthusiasmus*, an affect, is a sign of something very good: it signals that an agent's reason is successfully representing moral requirements to her. But Kant has nothing but disdain for *Schwärmerei*. He always uses *Schwärmerei* as a pejorative term for cases of "overstepping the bounds of human reason" (K2 5:85-86) – cases, for example, of wild metaphysical speculation about the nature of freedom or God (Review of Schulz 8:13, K2 5:85-86, K2 5:123, K3 5:275).

In the case of *Enthusiasmus*, the positive term, "reason is the cause of an affect that has the good as its object, and reason always handles the reins" (A 7:254). Kant is careful to emphasize here that when reason causes an affect, "our enthusiasm [*Enthusiasmus*] must be attributed to the *faculty of desire* [*Begehrungsvermögen*] and not . . . to a stronger sensuous *feeling*" (A 7:254) – a claim that fits the occurrence of reason-caused affects in Figure 3. As Kant says later, it is "the concept of freedom under moral laws that arouses an affect, which is called enthusiasm [*Enthusiasmus*]" (A 7:269).

What is the moral status of enthusiasm? The answer is somewhat complex. *Enthusiasmus* is a natural outgrowth (and an indication of success) of free, active moral willing. At the same time, Kant thinks that affects can be dangerous to freedom – so much so that "we would be ill advised deliberately to let [any affect] spring up in us" (A 7:253). So on the one hand, enthusiasm is a natural by-product of reason's representation of the moral law; on the other hand, affects impede reason. Kant's position seems to be this: the emotional life of the truly moral person may include so much reason-generated positive emotion (enthusiasm) that he or she will always have to be on guard against a corresponding loss of freedom (MM 4:408-409, K3 5:274-275).

A second reason-caused affect Kant describes as a kind of *astonishment* or *admiration* [*Bewunderung*]. There is a certain excitation of feeling, Kant says, that exists when a train of thought is unpleasantly stopped, then followed with an influx of unexpected thoughts. Someone inexperienced in the world – someone surprised and amazed at everything he sees – feels a naïve form of this affect (A 7:261). But there is also a non-naïve form of astonishment that occurs if one follows up surprise and amazement about nature with a certain sort of mental activity. In this case, astonishment can become a genuinely reason-produced affect: "If we follow up thoughtfully, with searching gaze, the order of nature in its great variety, we fall into amazement at a wisdom we did not expect – into admiration [*Bewunderung*] we cannot tear ourselves away from (we cannot be astonished [*verwundern*] enough). But then this affect is stirred up only by reason, and is a kind of holy thrill at seeing the abyss of the supersensible opening at our feet" (A 7:261). Note that unlike certain other affects, "astonishment/admiration" can be produced *only* by reason. Emotions like this that lead to awareness of the supersensible play an important role in the third *Critique*, as I will argue below.

The third reason-caused affect is a sort of courage Kant calls *fortitude*: "Courage as an affect . . . can also be aroused by reason and, accordingly, be true fortitude" (A 7:257). Kant has in mind here not "courage" understood as a virtue or disposition, but rather as that emotional force and "resoluteness" that accompanies courageous acts dictated by reason (A 7:257). "Here reason gives the determined man strength that nature sometimes denies him," Kant says (A 7:256-257). The existence of the reason-produced affect of fortitude has an interesting entailment. Recall that affects impede reflection for the short period that they exist; so even these three *reason-produced* affects result from reflection, but at least temporarily impede *further* reflection. In other words, Kant seems to think that reason can, for certain good reasons,

decide to let itself be impeded for a short time. The reason-produced affect of fortitude suggests an example: reason may see that to do one's duty, one needs a strong (if brief) emotional agitation – for instance, the pluck to run into a burning building to save a child, "to not shrink even from losing one's life in doing what duty commands" (A 7:259). Reason not only produces feelings and desires, but sometimes *must* produce *strong, reflection-inhibiting* emotions in order for an agent to do his or her duty. If "enthusiasm" is a common, possibly worrisome natural by-product of practical reason, "fortitude" is an affect that is *necessary* for the performance of some moral duties. This is a striking position given the image of Kant many readers take from the *Groundwork* and the second *Critique*. Perhaps all affects do "momentary damage to freedom and self-mastery" (A 7:267); the interesting point is that for Kant, it can be one's moral *duty* to briefly damage this very freedom and self-mastery.

III Reason-Caused Affects in the Third *Critique*

So far, the evidence that Kant thinks reason can produce (at least three) *affects* has come from the *Anthropology*, the last long work he published. Is there any reason to think Kant believed in reason-produced affects any *earlier*? The 1790 third *Critique* indicates that he did. Overlooked by many Kant interpreters is the position in the *Critique of Judgment* that reason can produce at least two of the three affects indicated above: "enthusiasm" and something very much like "astonishment/admiration."

Take the case for "enthusiasm" first. As in the *Anthropology*, "enthusiasm"¹⁹ is "an affect" (K3 5:275). "If the idea of the good is accompanied by affect, this is called *enthusiasm* [*Enthusiasmus*]," Kant says (K3 5:272). Also consistent with the *Anthropology* is Kant's claim in the third *Critique* that enthusiasm is a natural outcome of reason under moral laws, and that

this affect must be kept in check by reason. The passage in which this claim occurs is one of the most striking passages about morality and emotion in Kant's critical period:

It is indeed a mistake to worry that depriving this presentation [that is, the presentation of the moral law] of whatever could commend it to the senses will result in its carrying with it no more than a cold and lifeless approval without any moving force or emotion [*Rührung*]. It is exactly the other way round. For once the senses no longer see anything before them, while yet the unmistakable and indelible idea of morality remains, one would sooner need to temper the momentum of an unbounded imagination so as to keep it from rising to the level of enthusiasm [*Enthusiasm*], than to seek to support these ideas with images and childish devices for fear that they would otherwise be powerless. (K3 5:274)

Note that enthusiasm is an affect that naturally tends to arise from reasoning about morality.

This shows that Kant said the same things in 1790 about enthusiasm that he said in 1798. But more broadly, the passage also helps debunk the stereotype of Kant as a proponent of a detached, emotionless morality. For Kant, the agent who engages in moral reason is more likely to have trouble with an *excess* of emotional motivation than a *deficiency*.²⁰

"Astonishment" or "admiration" [*Bewunderung*], a second reason-produced affect from the *Anthropology*, also appears in the third *Critique*.²¹ There is a certain "aesthetic sublimity" to reason-caused enthusiasm, Kant says; but someone who is without affects is aesthetically sublime "in a far superior way." Such a state of mind – a state without affects – can somewhat paradoxically *arouse* an affect: "admiration [*Bewunderung*]," "an amazement that does not cease once the novelty is gone" (K3 5:272). In other words, reason can *produce* an affect in attending to the *absence* of affect. In the *Anthropology*, *Bewunderung* is the "holy thrill at seeing the abyss of the supersensible opening at our feet" (A 7:261); here, the holy thrill has a more specific object: the supersensible that opens to us is our own power to be without reason-impeding affects. Kant thinks that an awareness of this power to not be at the whim of empirical desire is

essential to morality; and since the (reason-caused) emotion of astonishment provides this awareness, it is clear that emotions play a positive role in Kant's ethics.

The experience of the sublime, developed at some length in the third *Critique*, plays a role that parallels the more specific affect *Bewunderung*. The experience of the sublime begins with an encounter with either absolute magnitude or with absolute power. The notion of infinity in mathematics is an example of absolute magnitude: numbers march on forever one by one, yet we seem to be able to wrap our arms around the notion of infinity as a totality. Forces of nature such as "hurricanes with all the devastation they leave behind, the boundless ocean heaved up, the high waterfall of a mighty river" (K3 5:261) are examples of absolute power: these forces threaten to destroy the agent, yet if she stands in a safe place, she discovers in herself an ability even greater than these natural forces – they cannot dominate her so as to force her to surrender her highest principles (K3 5:262). This experience of the sublime "reveals in us at the same time an ability to judge ourselves independent of nature, and reveals in us a superiority over nature that is the basis of a self-preservation quite different in kind from the one that can be assailed and endangered by nature outside us" (K3 5:261). Through either of these experiences, the agent notices in herself the existence of an "unlimited ability," something independent of nature (K3 5:259). This discovery is not immediately pleasurable. In the case of absolute magnitude, the agent feels frustration at the inability of her imagination to comprehend numerical infinity; in the case of absolute power, she feels the fear of imminent destruction by the forces of nature. Yet in both cases, this displeasure alternates with pleasure, since the discovery of the unlimited ability within herself is a pleasurable experience. This is a "pleasure that is possible only by means of a displeasure" (K3 5:260). Kant calls this alternation between displeasure and pleasure an "agitation" and a "vibration" (K3 5:260); the net effect he calls a "negative pleasure" (K3 5:245).

These descriptions of the feelings associated with the sublime place it clearly in the class of emotions called "stirrings" [*Rührungen*] above.²²

Defenders of the importance of the emotions argue that emotions are *morally informative*: emotions alert us to morally salient features about specific circumstances and objects of intrinsic value in those circumstances.²³ I note the presence of the sublime and *Bewunderung* in the third *Critique* because they show that emotions are morally informative in at least one sense for Kant, too: emotions can reveal our supersensible vocation as moral beings – a fact that Kant considers deeply important to human morality.

The third reason-produced affect of the *Anthropology*, *fortitude*, does not appear in the third *Critique*. But it is clear that Kant already has an impressive role for affects in morality as early as 1790. Affects play still other additional positive (but less directly moral) roles in both the third *Critique* and the *Anthropology*. For instance, occasional experiences of affects (such as laughter) are good for an agent's physical health (K3 5:332-335; A 7:262-263, 7:277-279). And Kant approves of grief as long as it is a "*vigorous* affect" that "has its basis in moral ideas" (K3 5:276).

By 1790, then, Kant seems to have a rich taxonomy of desires and feelings – some incompatible with moral reasoning (passions), some compatible with it under certain conditions (inclinations), and some actual *products* of it (desires such as "moral feeling" and certain stirrings and affects). As I have tried to show, all the important classes in Kant's taxonomy of desires and feelings are present by at least as early as 1790. Clearly present in the third *Critique* are the necessity of reason's determining role in morality (K3 5:269), the distinction between feeling and desire (K3 5:198, 5:187, 5:221-222), the existence of "moral feeling" or "respect," a specific desire produced by reason (5:222, 5:289, 5:292, 5:300), the susceptibility to sensible

pleasure and pain as a condition for experience (K3 5:187), the susceptibility to moral feeling as a condition for morality (K3 5:265), affects and passions and their nature (K3 5:272-273, esp. 5:272n), the existence of *reason-produced* affects (K3 5:272-275), and finally – something particularly emphasized in the third *Critique* – the existence of "stirrings" [*Rührungen*] like the experience of the sublime (K3 5:222, 5:226, 5:245, 5:258, 5:273). Kant mentions certain important affects (such as "fortitude") only after 1790; but the structure of the taxonomy is in place in the critical period.

For Kant, then, emotions play necessary, positive roles in morality. A clear taxonomy of Kant's emotion terms helps us understand what those roles are. A taxonomy also helps us avoid famous *misunderstandings* about Kant – for instance, the view that since Kant thinks lowly of inclinations, he must think lowly of all feelings and emotions. The positive roles for emotion are many. We have seen that the capacity for "moral feeling" is a necessary precondition for any moral action whatsoever. As the third *Critique* makes clear, no individual can be moral without a susceptibility to certain emotions (K3 5:265). In these ways, emotions are intrinsic elements of human morality for Kant. Further, various *reason-caused affects* can make moral contributions. Kant believes that there are cases where affects such as fortitude are necessary for the performance of certain moral duties. Reason must sometimes produce this affect if the agent is to perform certain demanding acts at all. In the "gradual reform of sensibility" necessary for becoming a good human being (R 47-48), feelings and emotions are crucial. A correct taxonomy helps us see why that is the case.²⁴

¹ See Marcia Baron, *Kantian Ethics Almost without Apology* (Ithaca: Cornell University Press, 1995), Barbara Herman, *The Practice of Moral Judgment* (Harvard: Harvard University Press, 1993), Nancy Sherman, *Making a Necessity of Virtue* (New York: Cambridge University Press, 1997), Paul Guyer, *Kant and the Experience of Freedom* (New York: Cambridge University Press, 1993) and *Kant on Freedom, Law, and Happiness* (New York: Cambridge University Press, 2000), and Allen Wood, *Kant's Ethical Thought* (New York: Cambridge University Press, 1999).

I use the following citation style for Kant's works within the text above: ([Book name] [*Akademie* volume]:[page number]). I use the following abbreviations for book names. Also listed are the translations I use. When I alter these translations for terminological consistency, I will so indicate.

- G *Groundwork of the Metaphysics of Morals* (1785) [Mary Gregor, trans., in Kant, *Practical Philosophy* (Cambridge: Cambridge University Press, 1996)]
- K2 *Critique of Practical Reason* (1788) [Mary Gregor, trans., in Kant, *Practical Philosophy* (Cambridge: Cambridge University Press, 1996)]
- K3 *Critique of Judgment* (1790) [Werner Pluhar, trans. (Indianapolis: Hackett, 1987)]
- MM *Metaphysics of Morals* (1797) [Mary Gregor, trans., in Kant, *Practical Philosophy* (Cambridge: Cambridge University Press, 1996)]
- A *Anthropology from a Pragmatic Standpoint* (1798) [Mary Gregor, trans. (The Hague: Martinus Nijhoff, 1974)]
- R *Religion within the Boundaries of Mere Reason* (1793) [George di Giovanni, trans., in Kant, *Religion and Rational Theology* (Cambridge: Cambridge University Press, 1996)]

² I will discuss three of these reason-caused affects later. Two of these affects appear in the *Critique of Judgment*; a third and very interesting affect ("fortitude") does not appear there. It is the existence of the class of reason-caused affects within Kant's taxonomy of the emotions that I mean is clearly in place in the 1790 third *Critique*.

³ For example, Werner Pluhar's translation of the third *Critique* uses "emotion" for *Rührung*.

⁴ For one particularly clear use of "inclination" as a technical term; see MM 6:213. Kant concedes here that "ordinary speech" does not use the term "inclination" as narrowly as he does. He suggests that if we want to accommodate ordinary speech, we *could* call "habitual desire from a pure interest of reason" (or "moral feeling") "*sense-free inclination*." In this paper, I will follow Kant's standard, narrower use of inclination as a technical term for only "habitual sensible desire."

⁵ *Lectures on Ethics*, Infield translation; quoted in Guyer, *Kant and the Experience of Freedom*, p. 339. Kant's distinction here parallels Hutcheson's distinction between justifying reasons and motivating reasons. (I thank Allen Wood for pointing out this parallel.) As I indicate above, Kant does not think that empirical desires or inclinations can play either role in an action that is done from duty.

⁶ However, see G 4:400.

⁷ Paul Guyer defends the importance of the active/passive distinction for Kant in *Kant and the Experience of Freedom* pp. 344-350.

⁸ Throughout this paper, I will be assuming that inclinations never have moral worth for Kant. This is a tempting assumption, given certain well-known passages (e.g., G 4:428). Some interpreters disagree: for instance, Paul Guyer has recently argued that inclinations can have moral worth when they fall under an agent's commitment to the fundamental maxim or morality – the maxim to always do what duty requires from respect for duty (Guyer, "Moral Worth, Virtue, and Merit," in *Kant on Freedom, Law, and Happiness*). My argument concedes more than this to a possible objector: suppose inclinations *do not* have moral worth for Kant; even so, I argue, he believes certain feelings and emotions have a positive moral status.

⁹ Kant generally uses the terms *Begehren* and *Begehrungsvermögen* for the broadest sense of desire in this picture, but sometimes he also uses *Begierde*. When Kant uses *Begierde*, he is often careful to indicate whether he is talking about the faculty of desire as a whole or the subclass of pleasure-caused desire. For example, at A 7:251 he calls pleasure-caused desire "desire [*Begierde*] in the narrow sense."

¹⁰ For example, early in the third *Critique*, Kant distinguishes *objective* sensation [*objectiv Empfindung*], such as the "green color of meadows," from *subjective* sensation [*subjectiv Empfindung*], such as "the color's agreeableness."

One more note about this passage: Kant proposes that for clarity we call only the latter (*subjective* sensation) "feeling" [*Gefühl*] (K3, 5:206). But Kant does not seem to follow this stipulation in the third *Critique* himself (see Werner Pluhar's footnote in his translation of *Critique of Judgment*, p. 48 (*Akademie* 5:206)). With this in mind, I treat *Gefühl* and *Empfindung* as more or less synonymous for the purposes of this paper.

¹¹ These are Kant's italics.

¹² I have placed "respect" in two places in Figure 1. Respect is a feeling, but since it is a feeling necessarily connected with a certain kind of reason-caused *desire*, and because Kant refers to some desires by the name of the feelings necessarily associated with them, I have represented it among these desires as well. Another note: recall that Kant seems to use the terms "respect" and "moral feeling" more or less synonymously until the *Metaphysics of Morals* in 1797 (see MM 6:399-403), when he distinguishes them as I explain below. The terms seem to be synonymous in the third *Critique*: cf., 5:222, 5:289, 5:292, etc.).

¹³ For consistency here, I substitute "sensible desire" for Gregor's translation, "sensuous appetite."

¹⁴ Kant even cautions against habitual *moral* maxims, since in this case an agent would also not be acting freely (MM 6:409).

¹⁵ Werner Pluhar, whose translation of the third *Critique* I follow in this paper, consistently renders *Rührungen* as "emotions." Both because "emotions" fails to fully capture the movement or alternation in feeling suggested by the German word *Rührungen* (and related words such as *rühren*, "to stir or move"), and because I am already using the term "emotions" in a more general way, I will change Pluhar's translations accordingly below.

¹⁶ *Affekt* is one of the least consistently rendered words in English translations of Kant. Mary Gregor often translates *Affekt* as "emotional agitation" in her 1974 translation of the *Anthropology* and her 1971 translation of the *Metaphysics of Morals*. Once Kant formally defines the term at A 7:251, she sometimes (but not consistently) renders the term as the English word "affect." However, in her 1996 Cambridge translation of the *Metaphysics of Morals* she generally chooses to use "emotion." Dowdell's 1978 translation of the *Anthropology* prefers "emotion." Again, for consistency, I have altered the translations included in this paper so that *Affekt* appears as the English word "affect."

¹⁷ For simplicity, I have drawn Figure 2 so that all affects are stirrings. This seems reasonable to me, but Kant does not settle this issue definitively. Affects clearly involve a sudden change of feeling, but it is unclear whether Kant thinks they *always* include a movement between opposite feelings. Nothing important that I say above depends on this issue.

¹⁸ Pluhar and Bernard render *Schwärmerei* as "fanaticism" in their respective translations of the third *Critique*. This rendering has the virtue of lessening the chance that English readers will confuse the term with *Enthusiasmus*. But an interesting historical point is preserved by rendering *Schwärmerei* as 'enthusiasm': Kant apparently intended *Schwärmerei* as a German translation of the English term 'enthusiasm' as *Locke* used it in the *Essay* (see *Locke's Essay*, Book IV, Ch. XIX). I thank Allen Wood for this point about translation.

¹⁹ As above, Kant's German word for "enthusiasm" in the passages quoted in this paragraph is *Enthusiasm* or *Enthusiasmus*, not *Schwärmerei*.

²⁰ Kant carefully proceeds to note that while the presentation of the moral law can naturally create "enthusiasm" [*Enthusiasm*], there is no danger of it creating "fanaticism" (*Schwärmerei*), a very different sort of emotion – one that includes the expectation of being able to sense something beyond the bounds of sensibility (K3 5:275). See my earlier comments on *Enthusiasmus* and *Schwärmerei* above.

²¹ Interestingly, Kant's discussion of *Bewunderung* occurs in the same section where he discusses the affect of enthusiasm (K3 5:272).

²² Does Kant think that the feeling associated with the sublime is necessarily *reason-produced*? This is not entirely clear in the third *Critique*. It does not need to be in order to play the positive role of making us aware of our noumenal freedom. However, Kant thinks that the sort of “astonishment” that results in awareness of the supersensible – an affect that appears related to the feeling of the sublime – can be reason-produced (see A 7:261).

²³ For more on the emotions as morally informative in Kant, see Nancy Sherman, *Making a Necessity of Virtue*, and “Reasons and Feelings in Kantian Morality,” *Philosophy and Phenomenological Research* 55 (June 1995): 369-377; see also Paul Guyer’s response to Sherman: “Moral Anthropology in Kant’s Aesthetics and Ethics: A Reply to Ameriks and Sherman,” *Philosophy and Phenomenological Research* 55 (June 1995): 379-391.

²⁴ I wish to thank Allen Wood for his generous comments on this paper. I also wish to thank Pierre Keller, Karsten Harries, the editor, and two anonymous referees for helpful comments on an earlier version of the paper.