

## Chapter 3

### *Two Concepts of Universality in Kant's Moral Theory*

Patrick R. Frierson

Kant's *Groundwork for the Metaphysics of Morals* (1785) famously emphasizes the *universality* of moral principles. His "categorical imperative" identifies universality as the fundamental criterion of duty (G 4:402, 421), and universality is the fundamental characteristic that "everyone must grant" distinguishes truly moral laws from mere practical rules (G 4:389). More than twenty years before writing *Groundwork*, Kant's *Observations on the Feeling of the Beautiful and Sublime* and *Inquiry concerning the distinctness of the principles of natural theology and morality* (1764) already laid out a moral theory stressing universality. In *Observations*, Kant writes that "true virtue can only be grafted upon principles and . . . will become the more sublime and noble the more universal [*allgemeiner*] they are" (Beo 2:217). Such virtue consists of "universal affection" and "universal respect," within which "one subordinates one's own particular inclination to such an enlarged one" that a perfectly virtuous "human being would . . . love and value . . . himself . . . only insofar as he is one among all to whom his widespread and noble feeling extends itself" (Beo 2:217). In the *Inquiry*, Kant's emphasis is on a "supreme universal formula" (UD 2:300) of morality and obligation that is "immediately necessary" (UD 2:298).

Despite their common focus on "universality," however, universality in Kant's early ethics differs from that of *Groundwork* in several fundamental respects.<sup>1</sup> This paper focuses on one key difference: In *Observations*, universality involves expanding the scope of benevolence, treating the welfare of *all* others as an end. This universality is

“teleological” or “objective” because it involves universalizing *objects* or *ends* of volition. Universality in *Groundwork* is quite different, emphasizing that the *grounds* for one’s actions must be possible grounds for the actions of all. This universality is “deliberative” or “subjective” because it emphasizes universalizing the point of view of the deliberating subject. Strikingly, Kant’s *Remarks*, written almost immediately after the publication of *Observations* and *Inquiry*, already show the beginning of Kant’s shift towards this more subjective universality. After laying out the difference between the objective universality of *Observations* and the subjective universality of *Remarks*, this paper discusses both philosophical and historical reasons for Kant’s shift.

### ***I. Universality in Observations***

*Observations* is not primarily ethics and certainly not the a priori “pure moral philosophy” of Kant’s *Groundwork*, but it includes a chapter devoted to “qualities of the sublime and the beautiful in human beings in general” (Beo 2:211), within which the sublimity of “true virtue” looms large. Kant emphasizes the importance of basing actions on *principles* (Beo 2:217-8), but not all principles are virtuous, only “universal rules” (Beo 2:215). Kant summarizes his account, saying that virtue “is the *feeling of the beauty and the dignity of human nature*” (Beo 2:217), and immediately unpacks this general statement in terms of *universality*: “The first is a ground of universal affection, the second of universal respect” (Beo 2:217). Thus “true virtue . . . become[s] the more sublime and noble the more universal [its principles] are” (Beo 2:217).

What sort of universality are principles supposed to have? Do they apply consistently across situations, or hold for everyone, or promote universal goods? One

apparent answer is the proto-utilitarian statement with which Kant summarizes universal affection and respect:

if this feeling had the greatest perfection in any human heart then this human being would certainly love and value even himself, but only in so far as he is one among all to whom his widespread and noble feeling extends itself. Only when one subordinates one's own particular inclination to such an enlarged one can our kindly drives be proportionately applied and bring about the noble attitude that is the beauty of virtue. (Beo 2:217)

Even this passage is ambiguous. If “loving” and “valuing” are akin to “respect for humanity” in *Groundwork*, “valuing” another might involve acting on principles that the other could share. Or does Kant have in mind benevolent concern for others, such that “universality” requires broadly extending benevolence?

Kant’s reference “kindly drives. . . proportionately applied” offers a clue. Virtue involves two universal feelings – affection and respect – that echo discussions earlier in *Observations* of sympathy and complaisance, which Kant describes as “kindly passion[s]” or “kindly feeling[s]” (Beo 2:216). These kindly feelings “cannot genuinely be counted as part of the virtuous disposition” though they are “good moral qualities that . . . to the extent that they harmonize with virtue, may also be regarded as noble” (Beo 2:215). With sympathy,

[S]uppose this sentiment moves you to help someone in need . . . , but you are indebted to someone else and by this means you make it impossible for yourself to fulfill the strict duty of justice; then obviously the action cannot arise from any

virtuous resolution, for that could not possibly entice you into sacrificing a higher obligation to this blind enchantment. (Beo 2:216)

Similarly, complaisance “is . . . far from being a virtue” because “From affectionate complaisance [one may] be a liar, an idler, a drunkard, etc.” (Beo 2:216-7). Sympathy and complaisance fail to be true virtue because they “only contingently agree with . . . universal rules of virtue” (Beo 2:215).

In discussing why kindly drives fail to be virtues, Kant shows how virtue can arise when these “feeling[s are] raised to [their] proper universality” (Beo 2:216).

Immediately after discussing how sympathy can go awry, Kant adds:

If . . . universal affection towards humankind has become your principle, to which you always subject your actions, then your love towards the one in need remains, but it is now, from a higher standpoint, placed in its proper relationship to your duty as a whole. The universal affection is a ground for participating in his ill-fortune, but at the same time it is also a ground of justice, in accordance with whose precept you must now forbear [helping another in a way that makes it impossible to repay one’s debts]. (Beo 2:216)

When sympathy is *universal*, it is a “ground of justice” rather than a “blind enchantment” away from justice. Even here, Kant does not explicitly say *how* universal sympathy grounds justice. At this time, Kant even seems uncertain of precisely how universalizing sympathy generates duties of justice, noting in *Inquiry* that “The ultimate fundamental concepts of obligation need to be determined more reliably” (UD 2:300). Kant is following up on an important “starting point” (UD 2:300) rather than laying out a fully developed moral theory.

Several factors, however, suggest an objective sense of universality in *Observations*. In *Inquiry*, Kant refers to “Hutcheson and others” as having “provided us with a starting point from which to develop some excellent observations” on moral principles (UD 2:300). Unsurprisingly, Kant’s own observations regarding universal affection reflect Hutcheson’s discussion of “calm settled universal benevolence” that “govern[s] and control[s] our particular generous as well as selfish affections.”<sup>2</sup> Even Kant’s distinction between “sympathy” and “complaisance” may draw from Hutcheson’s similar distinction between “love of complacence . . . and love of benevolence.”<sup>3</sup> Kant was attracted to Hutcheson for his emphasis on moral sense but also drew from him for substantive observations about moral feelings. Hutcheson’s approach to universality thus provides some guidance for understanding Kant’s. And in Hutcheson, “Benevolence is the foundation of all apprehended excellence in . . . Virtues.”<sup>4</sup> Like Kant, Hutcheson recognizes that particular acts of benevolence can conflict with what is really good and thus appeals to universalization. Moral appraisal is based on “influence of the action upon the *universal* natural good of Mankind . . . That which produces more [natural] good [i.e., happiness] than evil *in the whole* is acknowledged Good . . . [O]ur moral sense . . . recommend[s actions that] appear to have the *most universal* unlimited tendency to the *greatest and most extensive Happiness* of all rational Agents.”<sup>5</sup> Given Kant’s avowed enthusiasm for Hutcheson and the close similarity between his language and Hutcheson’s, it is reasonable to think that when Kant talks about making one’s affection more “universal,” he has in mind precisely the sort of objective universality that Hutcheson emphasizes, that is, an extension of one’s interest in others’ well-being to include *all* others.

Unsurprisingly, then, notes from ethics lectures Kant delivered in 1762-4 show evidence of his attempt to work out morality in terms of benevolence. Kant claims that God “gave us a disinterested feeling for the welfare, etc. of another,” argues that “the putting of oneself in the other’s shoes is . . . merely a means to vivacity, which presupposes the disinterested feeling” (*Herder*, 27:3), and insists that “morally good actions have a goodness that is assessed not by the effect but by the . . . intent” (27:4) and “the nature of the end determines the morality” (27:38). The *Remarks* themselves, as we will see, experiment with universality in terms of generalizing benevolence. And *Observations* emphasizes a sympathy closely aligned with benevolence, “a kindly participation in the fate of other people” that “moves you to help someone in need” (Beo 2:215-6). Insofar as sympathy helps interpret the “universal affection” of virtue, moral universality refers to the scope of beneficent regard, rather than a possibility for (all) others share one’s practical principles. An important footnote in *Observations* drives home this point:

sympathy . . . does not have in itself the dignity of virtue [because] . . . a suffering child . . . may fill our heart with this melancholy, while at the same time we may coldly receive the news of a great battle in which . . . a considerable part of humankind must innocently suffer dreadful evils . . . There is here no proportion in the effect at all, so how can one say that the universal love of mankind is the cause? (Beo 2:216n)

Unlike sympathetic concern for particular others, true virtue involves universal love that extends concern to *all* others (in due proportions).

One might expect a different sort of “universalism” in Kant’s reference to universal *respect* and its related “kindly feeling,” complaisance. But complaisance, which is ultimately oriented towards “mak[ing] ourselves agreeable to others through friendliness” (Beo 2:216), is a “charming” trait more akin to the “agreeable” virtues of Hume’s *Enquiry* than to Kant’s *Groundwork*. Moreover, Kant’s attitude towards complaisance is considerably more dismissive than his attitude towards sympathy. (This dismissive attitude culminates in the *Remarks*, which do not discuss complaisance [*Gefälligkeit*] at all.) Finally, while the structure of *Observations* suggests a link between respect for humanity and properly universalized complaisance, Kant’s explanation of how complaisance is improved involves not universalizing it but bringing it “in[to] accordance with the rules for good conduct in general” through combining it with “self-control and . . . principles” (Beo 2:216-7). Universality does not enter significantly into Kant’s corrections for complaisance, and complaisance, accordingly, offers very little guidance for understanding the universality of virtue in *Observations*.

In *Observations*, then, true virtue requires acting on principles that are “universal” in the sense that they extend concern for others’ welfare and wishes to cover *all* others. This universality is broadly utilitarian (like Hutcheson’s), in that it values the goods of all human beings “proportionately,” such that one will “love and value” each (including oneself) “only . . . as . . . one among all”.<sup>6</sup>

## ***II. Universality in the Remarks***

Kant’s continued interest in moral theory is evident in the “Remarks” written in his personal copy of *Observations*. Many of these cohere with the objective universality

of *Observations*. Kant discusses a “universal love of humankind” (Bem 20:25) and explains that the “will . . . is good for itself if it wants everything that contributes to its perfection (pleasure), and good for the whole if at the same time it desires the perfection [pleasure] of all (Bem Bem 20:138). Moral goodness requires pursuing others interests as one pursues one’s own. And when Kant considers ways in “the sympathetic sentiment” can be “universal,” he emphasize the scope of one’s “altruism,” the “help” one offers (Bem Bem 20:173). Even with respect to lying, Kant’s reflections reinforce objective universality.

Because in society all mine and thine depends on contracts, yet these [depend] on keeping one’s word, love of truth is the foundation of all social virtue, and lying is the main vice against others . . . (Bem 20:153)

The problem is not that lying contradicts itself when (subjectively) universalized, but that lying undermines social virtue and commerce. In these and similar remarks, Kant retains and develops objective universality of *Observations*.

But elsewhere, Kant experiments with new ways of thinking about universality. For example, Kant turns from the “social virtue” account of lying to one of “strict obligation” that specifically distinguishes such obligations from benefit to others.<sup>7</sup>

[T]ruthfulness does not depend on philanthropy, but on the sense of justice . . . This sense . . . has its origin in the nature of the human mind, through which one judges what is categorically good (not useful), not according to private benefit or benefit to others, but through supposing the same action in others; if a contradiction and contrast then arises, it displeases; if harmony and unison arise, they please. (Bem 20:157)



Anticipating the subjective universality of his mature moral philosophy, Kant connects this universality with the “categorical” nature of morality and argues against “philanthropy” as the basis for morals. By describing moral sense in terms of “contradiction” when “supposing the same action in others,” Kant provides a way of interpreting universality without universal *benevolence*. The reference to “private benefit or benefit to others” seems to directly disallow Hutchesonian-Humean accounts of virtue according to which moral sense approves either of actions oriented towards public rather than private benefit (Hutcheson) or to private and public benefit (Hume). Rejecting both in favor of subjective rather than objective universality, Kant moves from his sentimentalist influences towards his mature moral philosophy.

### ***III. Why subjective universality?***

This section discusses three philosophical reasons for Kant’s turn from objective to subjective universality. Kant’s anti-consequentialism, desire for certainty in morals, and concerns about motivational efficacy combined to prompt his shift from objective to subjective universality.

Kant’s rejection of consequentialism as a basis for moral judgment is among the most well known philosophical claims from Kant’s *Groundwork*. After all, a “categorical” imperative is precisely one to be obeyed regardless of possible consequences. Even in *Inquiry*, Kant insists that the “immediate ugliness” of bad actions emerges as long as we “do not straightaway focus our attention on . . . consequences” (UD 2:300), and *Observations* ascribes a sublimity to virtue that is independent of its effects.

The fact that moral worth cannot be evaluated based on consequences of one's actions is often taken – arguably even by Kant (see *G* 4:399-400) – to imply something like subjective universalism in morals. But Kant's *Observations* (not to mention Hutcheson's and Hume's moral theories) shows that resistance to evaluating specific actions and attitudes based on their consequences is consistent with a broad consequentialism that takes *intended* or *customary* results of (types of) actions or attitudes as criteria of virtue. Even while insisting that virtue is sublime in itself, *Observations* advances a basically objective account of universality. And in *Remarks*, Kant combines the *consequentialist* claim that “[t]he . . . will . . . is good . . . if . . . it desires the [happiness] of all” with the *anti-consequentialist* claim,

However destitute the human being who has this will may be, the will is still good.

Other things may be useful; other human beings may do a lot of good . . . through a lot of power and a small degree of will; yet the ground of willing the good is still uniquely and solely moral. (Bem 20:138)

Similarly, Hutcheson insists that the moral sense *immediately* approves of virtue rather than approving of it for its effects, but nonetheless argues that what it is for an action to be virtuous is for it to promote (or aim to promote) the best consequences for the greatest number.

It is thus important to distinguish two types of anti-consequentialism. For the Kant of *Observations*, as well as Hutcheson, Hume, and most “consequentialist” moral philosophers today, moral worth of actions is determined by the extent to which those actions are oriented towards good results, primarily for others. This results-orientation is consequentialist, but such moral philosophers still can – and typically do – resist basing

moral evaluations on consequences *directly*. Acting out of concern for others' well-being is good even if it turns out badly, and unjustly promoting personal advantage is bad even if it happens to benefit all concerned. This "shallow" anti-consequentialism is anti-consequentialist at the level of immediate moral evaluation of actions but allows consequentialist reasoning at a deeper level of moral evaluation. But there is a second sort of anti-consequentialism – "deep" anti-consequentialism – that goes further. According to this view, which lies at the heart of Kant's *Groundwork*, "the purposes we may have for our actions . . . can give actions no unconditional and moral worth . . . [Moral] worth . . . can lie nowhere else *than in the principle of the will* without regard for the ends that can be brought about by such an action" (G 4:400). This view rejects not only the crude consequentialism that evaluates actions by their actual consequences but even the more plausible consequentialism that identifies the morally good feature(s) of motives as identical to or closely associated with interest in certain ends, such as the welfare of others.

Kant's *Remarks* show a dramatic increase in Kant's concern with avoiding consequentialism in ethics. These *Remarks* persistently distinguish what is useful from what is (morally) good (see Bem 20:24, 118, 133, 138, 146, 156-7, 168). And while some comments are consistent with shallow anti-consequentialism, others shift towards the deep anti-consequentialism that characterizes Kant's later moral theory. Kant's claims that "in moral matters, the noble much not be considered from the viewpoint of usefulness" (Bem 20:118) and that "utility . . . indicates no perfection complete in itself" (Bem 20:133) preclude utility even at the level of determining whether a motive or

disposition is virtuous, and Kant rejects the reduction of morals to “duty towards others” because this would undermine the “immediate ugl[iness]” of vice (Bem 20:24).

Shifting from shallow to deep consequentialism helps motivate a new conception of universality because deep anti-consequentialism is inconsistent with identifying virtue and universal benevolence. Benevolence necessarily aims at promoting others’ welfare, and even if it does not reduce morality to “the greatest good for the greatest number,” it emphasizes orientation towards consequences as proper bases for moral appraisal. Unlike objective universalism, the subjective universalism of *Groundwork* coheres with deep anti-consequentialism. Whether something can be willed by all similarly situated agents does not directly appeal to consequences at all,<sup>8</sup> and thus provides a deeply anti-consequentialist “groundwork” for ethics. However, in contrast to *Groundwork*’s suggestion that deep anti-consequentialism *implies* that morality is a matter of subjectively universalizable maxims (see G 4:400), *Remarks* explores several forms of deep anti-consequentialism, such as taking principles of justice as immediately given, or deriving them through “contradictions” that do not appeal to universality at all. To sustain deep anti-consequentialism, one needs only some way of evaluating practical principles *directly*, without appeal to actual *or even intended* consequences. To be deeply anti-consequentialist, a moral theory must ascribe to practical principles a worth that is not reducible to the consequences at which those principles aim or to which they typically lead.

Beyond deep anti-consequentialism, a second philosophical commitment – Kant’s concern with moral clarity – helps push towards subjective universality. Kant worries that “everything passes by us in a river,” laments “changeable taste,” and asks, “Where do I

find fixed points in nature that a person can never mistake and that could give him signs as to which bank he must head for?” (Bem 20:46). This desire for “fixed points” leads Kant to postulate a “certainty in ethical judgments . . . just as great as with logic” (Bem 20:49). In contrast to his insistence in *Inquiry* that the “feeling of the good” is “unanalyzable” (UD 2:299, see too PPH 27:5), Kant’s *Remarks* offer, “through analysis I will make it as certain to a human being that lying is repulsive as that a thinking body is incoherent” (Bem 20:49). To make ethical judgments as certain as logical ones, Kant turns to contradiction. Just as logical contradiction is an epistemic failing not justified by appeal to bad consequences (false beliefs), Kant considers the possibility of practical contradiction as a non-consequentialist approach to the immediate ugliness of morality (Bem 20:93).

In keeping with *Inquiry*’s emphasis on “indemonstrable material principle[s] of obligation” (UD 2:300), Kant’s first attempts to connect the immediate ugliness of morality with contradiction emphasize contradictions between practical principles and facts about the (human) world.

In subjection there is not only something externally dangerous but also a certain ugliness and a contradiction that . . . indicates its unlawfulness. An animal is not yet a complete being because it is not conscious of itself; and its drives and inclinations may be opposed by another or not . . . But that the human being himself should . . . need no soul and . . . have no will of his own, and that another soul should move my limbs is absurd and perverse. (Bem 20:93)

The “contradiction” is between positing that another human being has no will of its own by subjecting him to one’s will and the obvious fact that human beings, unlike animals,

*do* have wills of their own. Similarly for property rights, Kant insists that “the other . . . not call his own what I have worked upon, since otherwise he would presuppose that his will moved my body” (Bem 20:67). One makes something one’s own through “perform[ing] those actions that designate it as mine, cut[ting] down the tree, mill[ing] it, etc” (Bem 20:67). Subjecting objects to my will makes them mine. When another claims something *I* worked on, he implicitly asserts that his will subjected the object, hence that his will (not mine) cut down the tree, and thus that his will moved my body. This “contradicts” the fact that “[my] body is mine . . . and is moved by my faculty of choice” (Bem 20:66). These examples show how Kant’s attempts to avoid consequentialism<sup>9</sup> and ground morals as securely as logic lead him to a contradiction-based moral principle. But these examples also show that emphasizing contradiction need not imply subjective universality for morals. Instead, here Kant shifts away from universality altogether, emphasizing contradictions with essential truths about human nature or property-acquisition.

But even as Kant offers examples that replace universality with contradiction, he experiments with combining contradiction and a new, subjective universality. In discussing property, he writes,

A will that is to be good must not cancel itself out if . . . taken universally and reciprocally . . .<sup>10</sup> [W]hen a human being calls some things his own, he thereby *tacitly* promises in similar circumstances, through his will, not about something—  
(Bem 20:67)

Unfortunately, this remark breaks off here, without Kant finishing his exploration of a contradiction internal to the will itself, a contradiction rooted in a new sort of

universality. Kant has good reason to experiment with this conception of contradiction, since appealing to facts about property or human wills might provide models for moral “analysis,” but such analysis would be more akin to physics than to the logical certainty Kant sought (Bem 20:49). When Kant uses language of “contradiction” later in *Remarks*, he is clearer about what contradiction he looks for: one “suppose[s] the same action in others” and considers whether “a contradiction and contrast then arises” (Bem 20:157). Here Kant returns to private property and finishes his previously unfinished thoughts.

An action that contradicts itself, when considered from the perspective of the general will of human beings, is . . . morally impossible (forbidden). Suppose I were about to take the fruits of another. If I then see that, under the condition that what one acquires will soon be snatched away, nobody wants to acquire anything, then I will desire another’s goods from the private point of view while rejecting them from the public one. (Bem 20:161)

The private desire to take another’s goods contradicts the (subjectively) universalized generalization of one’s action. Kant finishes the argument begun many pages earlier in a way that combines strong commitment to avoiding consequentialism with his interest in the principle of contradiction as a quasi-logical standard of moral judgment through invoking a subjective universality test akin to that in his later *Groundwork*.<sup>11</sup>

Anti-consequentialism and interest in contradiction are not the only factors drawing Kant towards subjective universality. Another important consideration relates to the motivational efficacy of universal benevolence (or sympathy). Throughout *Remarks*, Kant attends to problems of moral motivation, especially how civilized human beings can overcome temptations to selfishness and actually do what is right.<sup>12</sup> Among the most

important moral-motivational problems arises for universal benevolence. As benevolence grows more universal, it grows less motivationally effective. With respect to sympathy, *Observations* notes,

[A]s soon as this feeling is raised to its proper universality, it is sublime, but also colder. For it is not possible that our bosom should swell with tenderness on behalf of every human being and swim in melancholy for everyone else's need. (Beo 2:216)

*Remarks* reemphasizes this point and directly applies it to the formulation of virtue as universal love proposed in *Observations* (Beo 2:217):

The universal love of mankind has something high and noble in it, but in a human being it is chimerical. If one aims for it, one gets used to deceiving oneself with longings and idle wishes. (Bem 20:25, see too Bem 20:45)

In *Remarks*, Kant further argues that universal love is not only weaker but also less effective than particular love.

Compassion is an *affect* of benevolence toward the needy, according to which we imagine that we would do what is in our power to help them; it is thus for the most part a chimera, because it is neither always in our power nor in our will. (Bem 20:135, see too Bem 20:173)

Not only is universal compassion not motivationally effective (not “in our will”), but even if it did motivate, we lack *power* to actually promote the welfare of all others.

One important aspect of Kant's multifaceted response to this motivational problem is his growing commitment to a moral motive independent of benevolence. Already in *Observations*, Kant seems to posit the “*feeling of the beauty and the dignity of human nature*” (Beo 2:217) as something that grounds but is not necessarily identical to



universal affection and respect. *Inquiry*'s description of vice's "immediate ugliness" also suggests a moral motive beyond benevolence. And *Remarks* makes explicit the need for "an immediate [inclination] to good actions" (Bem 20:18). Eventually, in his *Groundwork* and later moral philosophy, Kant unpacks this motive as "respect for the [moral] law" (G 4:400), and motivational issues with benevolence help show one reason for his later positing an independent moral motive. Kant's concern with the motivational efficacy of an objectively universal feeling of benevolence does not *require* a shift to subjective universality,<sup>13</sup> but appealing to immediate moral motives opens room for and encourages Kant to develop criteria of goodness not rooted in inclinations like benevolence or complaisance. And motivational problems with universal benevolence at least encourage shifting to subjective rather than objective universality as a moral standard.

We can thus highlight three factors that led Kant from objective to more subjective universality in moral theory. Kant's opposition to consequentialism grew during this period, making objective universality less acceptable to him. His interest in moral certainty led him to a quasi-logical principle of contradiction as a way of establishing moral rectitude, a standard that provided a way to make sense of subjective universality. And his concerns with the motivational efficacy of universal benevolence required and enabled him to look for an alternative to objective universality.

#### ***IV. Influences***

The last section highlighted three considerations internal to Kant's moral-philosophical commitments in the early 1760's that encouraged shifting from an objectively universal moral standard to a subjectively universal one.<sup>14</sup> But for Kant, the 1760's were not only a time of intense philosophical reflection, but also a time of extreme engagement with philosophical developments from around Europe. Kant emphasizes, in his *Inquiry* and "Announcement of . . . lectures for . . . 1765-1766," that the British sentimentalists were an important influence during this time (UD 2:300, 311).<sup>15</sup> And Kant's *Remarks*, where he began exploring subjective universality, is permeated with reflections on Rousseau. What influence might British sentimentalist and/or Rousseau have had on Kant's shift from objective to subjective universality?

At first, British sentimentalism seems an unlikely source for the shift. Hutcheson is a primary source for Kant's objective conception of universality, and Hume seems at least as interested in ends, even if not in universality. Hume's *Enquiry* divides natural virtues into those that are useful or agreeable to oneself or others and insists,

we must . . . conclude it impossible for . . . man to be totally indifferent to the well or ill-being of his fellow-creatures, and not . . . to pronounce . . . that what promotes their happiness is good, what tends to their misery is evil, without any farther regard or consideration.<sup>16</sup>

Hume seems as interested as Hutcheson in intended or customary effects of actions, and in that sense seems an unlikely source for non-objective moral standards.

Nonetheless, British sentimentalism offers at least two possible sources for a more subjective universality. First, while Hume's moral philosophy emphasizes the role of sympathy in extending interest to other *objects* (the pleasures and pains of others), Hume

also includes *subjective* generality (if not strict universality) in his moral theory. This argument is clearest in the *Treatise*, where Hume writes,

[T]o prevent . . . continual *contradictions*, . . . we fix on some . . . *general* points of view; and always, in our thoughts, place ourselves in them, whatever may be our present situation.<sup>17</sup>

These “points of view” involve *subjective* generality. We judge in ways that we imagine others will also judge. And although Hume uses subjective universality to promote a fundamentally objective criterion of praise and blame (“the general interests of the community”), he still introduces a kind of moral argument that might spark Kant’s interest in a new way of “generalizing” moral principles.

Whereas subjective universality would be a substantial inference from a relatively tangential part of Hume, it shows up quite directly in Adam Smith. In his *Theory of Moral Sentiments*, Smith distinguishes his account of sympathy from Hume’s in that, for Smith, true sympathy (where one fully shares another’s sentiments) is intrinsically pleasurable, even if the sentiments one shares are painful: the “emotion which arises from . . . observing the perfect coincidence between this sympathetic passion in himself, and the original passion in the person principally concerned . . . is always agreeable and delightful,” even if the original passion is an unpleasant one.<sup>18</sup> Given this non-Humean account of sympathy, Smithian moral approbation is based on the pleasure of sympathy itself rather than (as in Hume) sympathetically sharing in pleasurable effects. Thus for Hume, one approves of benevolence because one sympathetically shares the beneficiary’s pleasures. For Smith, one approves of benevolence because one pleasurable sympathizes with sentiments of the benefactor. And thus Smith can explain why one might approve of

an attitude like resentment (that is unpleasant to all affected by it), insofar as it is a sentiment with which one can sympathize. While Humean sympathy supports a moral theory emphasizing pleasures and pains caused by actions or dispositions, Smith's conception of sympathy leads to a moral theory emphasizing shareability of actions and attitudes: "To approve of the passions of another . . . is the same thing as to observe that we entirely sympathize with them" (TMS I.i.3.1, p. 16). Smithian sympathy – and thus Smith's standard for moral approbation – is *subjective*, while Humean sympathy – and thus Hume's standard – is basically *objective*. Reading Smith in relation to Hume and Hutcheson could certainly have prompted Kant to reflect on sympathy in a way that would naturally support a shift to subjective moral universality.

A key problem for identifying either Smith or Hume as a historical source for Kant's moral theorizing is that Kant could not read English and neither Hume's *Treatise* nor Smith's *Theory of Moral Sentiments* were available in German in the mid-1760's. Of course, there are ways that Kant *might* have accessed either text despite these problems. With respect to Hume, the textual issue is not as urgent because while the clearest allusions to subjective universality are in the *Treatise*, some allusion show up in the *Enquiry* as well.<sup>19</sup> Unfortunately subjective universality is much less prominent in Hume than in Smith. With respect to Smith, Kant could have discussed Smith with his friend Joseph Green, with whom he certainly discussed other developments in British philosophy (see Kuehn 2001, 272-3). Moreover, although there would be no *German* translation of *A Theory of Moral Sentiments* until 1770, there was a complete and reliable *French* translation of the entire work published in 1764 and publicized in *Correspondance litteraire* – a journal to which Kant had access – in 1765. Given Kant's

intense focus on both British sentimentalism and Rousseau's *Emile* and *Social Contract*, it would be reasonable for him to seek out a French edition of Smith's *Theory of Moral Sentiments*. And given that the *Remarks* in which notions of subjective universality are clearest come late, the date at which Kant might reasonably have perused the French translation of *Theory of Moral Sentiments* fits well with the onset of his interest in subjective universality.<sup>20</sup> Unfortunately, it is impossible to know with certainty many of the books that Kant read during this time. As J. G. Hasse notes, "[Kant's library] was not at all large, for he did not buy books readily, but rather had them sent to him unbound from the bookshop, read them, then sent them back" (Hasse 1804, p. 32). But Kant could read French, had French works in his personal library (see Warda, 1922), and would have had good reason to seek out Smith's work at this time.

Access to the French edition of *Theory of Moral Sentiments* would also explain a *Reflexion* in which Kant asks whether "sympathy with others or the impartial spectator" is the "rule of application" for moral principles. As Samuel Fleischacker has noted, this distinction "nicely encapsulates the contrast between what Smith and Hume . . . consider central to ethical judgment" (see Fleischacker 1991, p. 252). Because Smith uses the "impartial spectator" to highlight the subjective universality necessary for moral appraisal, this *Reflexion* highlights the concern with the relative importance of objective and subjective considerations in morals in a way that ties them directly to the difference between Hume and Smith. The dating of this *Reflexion* is uncertain. At the latest, it could have been written in 1770, but Berger conjectures that it could have been written as early as 1764. The later date would have just barely given Kant time to read the German translation of *Theory of Moral Sentiments* immediately after it came out. An earlier date

would put this *Reflexion* squarely in line with Kant's *Remarks*. In either case, access to the French edition would help explain how Kant could have formulated such a clear understanding of the contrast between Smith and Hume at a relatively early date.<sup>21</sup>

Beyond British sentimentalism, an obvious source for Kant's reflections on subjective universality is Rousseau, with whom Kant was actively engaged during this period.<sup>22</sup> The challenge is to explain what Kant might have discovered through sustained engagement with Rousseau's works – an engagement that began long before Kant's *Remarks* (see 2:247) – that would gradually lead to an interest in subjective universality. One reasonable conjecture is that Kant's attention increasingly shifts from the moral theory of *Julie* and *Emile* Books I-IV to Rousseau's political theory, as described in *Emile* Book V and especially the *Social Contract*.

Rousseau's most important explicit discussions of morals are in *Emile*, Book IV. The first is the "Profession of Faith of the Savoyard Vicar," where Rousseau's emphasis is on the immediacy of moral feeling, the moral roles of reason and sentiment, and fact that all people have similar moral convictions. But this "Profession" does not characterize conscience's content with anything like the sort of clarity that Kant, even in *Observations*, seeks. Earlier, however, in describing how to cultivate Emile's conscience, Rousseau characterizes how true morals emerge from natural pity and therein highlights universality:

To prevent pity from degenerating . . . , it must . . . be generalized and extended to the whole of mankind. Then one yields to it only insofar as it accords with justice,

because of all the virtues justice is the one that contributes most to the common good of men. (253)

Just as Kant's *Observations* generalizes sympathy, Rousseau universalizes pity, and as in Kant, this universalization is objective. Justice is defended in consequentialist terms as "contribut[ing] to the common good" and the discussion takes place in the context of generalizing Emile's self-interest and thereby making "his cares consecrated to the happiness of others" (252). When interest in others' happiness is universalized, "[i]t [will be] of little importance to him who gets a greater share of happiness provided that it contributes to the greatest happiness of all" (252-3). On initially reading Rousseau, Kant would have found an account of morals that fit well with his own Hutchesonian account. Where Hutcheson discusses "benevolence" and "sympathy," Rousseau emphasizes "pity," but both insist on universalizing concern for others' well-being.

But in *Emile's* concluding pages and especially in the *Social Contract*, Rousseau introduces rudiments of a new conception of universality. These works continue to make use of objective conceptions of universality to a considerable degree, discussing "public utility" or "common interest" and insisting that "the law can never have anything but a general *object*."<sup>23</sup> But they also emphasize *subjective* universality through the important concept of the "general will." Rousseau "sums up in a formula" the social contract: "*Each of us puts his goods, his person, his life, and all his power in common under the supreme direction of the general will.*"<sup>24</sup> For Rousseau, "What unites the general will is . . . [that] in this institution, everyone necessarily submits the conditions which he imposes on others."<sup>25</sup> For Rousseau, "the universality of the will and that of the object" are two distinct sorts of universality "combined" in the general will.<sup>26</sup> Because the general will

requires “universality of the will” – that is, the consent of all –it takes on a “universality of . . . the object” – that is, an application to all for the welfare of all. In the *Social Contract*, subjective universality of consent is conceptually prior to objective universality of public utility.

Importantly, Rousseau’s turn to the “general will” is tied to an emphasis on “freedom” and “self-governance.” Those submitting to the general will “obey no one, but only their own will,”<sup>27</sup> and only submission to the general will provides “moral freedom, which alone makes man truly the master of himself; for the impulsion of mere appetite is slavery, and obedience to the law one has prescribed to oneself is freedom.”<sup>28</sup> Strikingly, freedom plays virtually no role in Kant’s *Observations* but becomes a central concern of *Remarks*, and many of Kant’s allusions to subjective universality take place in the context of increased attention to freedom. One series of remarks is particularly striking. Kant first appeals to “moral feeling” to claim that “The will is perfect insofar as . . . it is the greatest ground of the good in general” (Bem 20:136). Kant adds that this will must act “in accordance with the laws of freedom” (Bem 20:136), but only a few pages later does he thematize freedom itself: “the greatest inner perfection . . . consist[s] in the subordination of all of the capacities and receptivities to the free faculty of choice” (Bem 20:146). And only here does Kant ground obligation in what “is necessary through the general will,” that is, what is necessary when “the human being considers himself at the same time in *consensus* with the general will” (Bem 20:146). Kant’s shift from objective to subjective universality may in part have been prompted by rereading Rousseau with a focus on freedom and thereby foregrounding Rousseau’s *political* theory. Kant thus transforms the



subjective universality prominent in Rousseau's political theory into a moral principle that can be deeply anti-consequentialist, certain, and motivationally effective.<sup>29</sup>

## V. Conclusion

This paper focused on a shift in Kant's conception of moral universality, a shift that occurred in 1764-5 in the light of philosophical commitments that Kant held even earlier and in response, perhaps, to new insights from Hume, Smith, and/or Rousseau. This shift eventually became an important part of the moral theory for which Kant became famous. The difference between these concepts of universality, and Kant's reasons for shifting his views, is important in several respects, and this conclusion can only briefly highlight some of them. For one, seeing the way in which Kant's view changed can help readers of both Kant's early and late moral theories avoid overly hasty (mis)interpretations of Kant's views. Kant's *Groundwork* cannot simply be summed up as promoting "universalization" in morals. And Kant's articulation of "respect for humanity" in *Groundwork* must be distinguished from mere concern for others' welfare. Improved understandings of Kant's texts can also help provide responses to common objections to Kant's moral philosophy. Recognizing that respect for the moral law grows out of Kant's solution to a motivational problem for *sentimentalist* moral theories can help show that respect is not a merely an ad hoc way of building motivational force into a fundamentally rationalist moral theory. And once subjective and objective universality are contrasted, we can better see how Kant's later moral theory might meet concerns about treating "special relations" such as friends and family in different ways than one treats others. While a theory of objective universality runs into problems with special

relations, Kant's subjective universality can deal with it straightforwardly in terms of the possibility for all to give special preferences to friends and family. Given that objectively universal love is chimerical, one even has good reason to encourage attention to particular others rather than the human species as a whole (see Bem 20: 173). Finally, seeing the origins of Kant's mature moral philosophy helps one better understand his moral priorities. Contrary to what one might think on reading *Groundwork*, opposition to consequentialism, desire for moral certainty, and concern about moral motivation were more fundamental to Kant than any particular formula of morality, and certainly more fundamental than any commitment to a purely rational foundation for morals.<sup>30</sup> While understanding the biographical origin of Kant's ideas cannot directly take the place of attending to his actual arguments, it can help direct attention to new ways of either defending or criticizing both the moral theories from which Kant made his start and those that he eventually developed.

---

<sup>1</sup> There are important differences between the universality of *Observations* and *Groundwork* on which I do not focus. For one, moral "universality" in *Observations* is still specifically *human*, rather than applicable to all "rational beings." (*Observations* is so far from claiming this radical universality that Kant even suggests the possibility of different ethical standards for men and women (Beo 2:228f.).) For another, *Observations* does not see "universality" as requiring an *a priori* basis for morality. And finally, while universality in Kant's *Groundwork* is tied to freedom and autonomy, neither of these concepts plays any significant role in *Observations* or *Inquiry*.

<sup>2</sup> Hutcheson 1755, p. 74.

<sup>3</sup> Hume 1978, p. 102.

<sup>4</sup> Hutcheson 2004, p. 118.

<sup>5</sup> Hutcheson 2004, pp. 118, 126.

<sup>6</sup> *Observations* 2:117. Cf. Hutcheson's *Inquiry* (Hutcheson 2004, pp. 122-3, 128-9). In describing this account of virtue as utilitarian, it is important to emphasize that the account is *rule*-utilitarian (focused on principles rather than specific acts) and also

---

*motive*-oriented rather than result-oriented, such that one is evaluated based on what one intends rather than what one's acts or even principles will actually (tend to) bring about. These features allow Kant to combine the objective universality of universal benevolence with an anti-consequentialist insistence upon the "immediate ugliness" (UD 2:300) of moral wrongdoing. (I discuss this in more detail in section 3.)

<sup>7</sup> Kant does not limit his new subjective test to the case of lying, but applies it to generosity (Bem 20:158) and private property (Bem 20:161).

<sup>8</sup> Consequences might still be relevant to assessing whether maxims could be willed universally, but consequences are neither *necessarily* nor *immediately* relevant.

<sup>9</sup> Neither slavery nor theft is wrong because of real, intended, or customary consequences. Even if (some kind of) slavery or theft were consistently beneficial overall and one enslaved or stole with benevolent intentions, there would still be a "contradiction that indicates their unlawfulness."

<sup>10</sup> Strikingly, the ellipses mark the deletion of the phrase quoted above: "for the sake of this, the other will not call his own what I have worked upon, since otherwise he would presuppose that his will moved my body." In the midst of his insight into the possibility of a formal moral principle based on subjective universality, Kant slips (back) into a realist appeal to substantive material practical principles. Perhaps that explains why the following sentence breaks off. Kant is on to something, but has not yet figured out how to articulate it consistently. As we will, Kant returns to this example later (Bem 20:161).

<sup>11</sup> The test is not exactly the same. Kant has not isolated "maxims" as relevant foci of moral evaluation and has not worked out the distinction between what have become known as "contradiction in conception" and "contradiction in will" (see e.g. Korsgaard 1996, pp. 78). Kant seems to find the contradiction between what one wants in this case and what all would want in a world where everyone acts as one plans to act. That is more complicated and less straightforward than his account in *Groundwork*, but it is a close approximation to it, and a clear step in the direction Kant eventually goes.

<sup>12</sup> E.g., Bem 20:15, 191. Kant thus considers the possibility that religion can be motivationally important for virtue (see Bem 20:12, 16-19, 57, 104), and Kant actively endorses the basically Rousseauian prescription one does better to "restrict one's opulent inclinations" than "by keeping them, [to] invent remedies against their insults" (Bem 20:16).

<sup>13</sup> Freedom of the will (emphasized in *Remarks*) might make moral motivation less urgent. And Kant also works on a careful articulation of a sort of universal sympathy that can be genuinely effective (Bem 20:173) that does not require subjective universality.

<sup>14</sup> Strikingly, all three factors were present in *Observations* and *Inquiry*. While Kant had not yet emphasized the *deep* anti-consequentialism he would introduce in *Remarks*, his *Inquiry* already emphasized vice's "immediate ugliness." And while *Inquiry* allows for "indemonstrable material principle[s] of obligation," Kant still aimed to "attain the highest degree of philosophical certainty in the fundamental principles of morality" (UD 2:300). And motivational problems with benevolence are, if anything, stated *more* dramatically (though with less nuance) in *Observations* than *Remarks*. While these three factors thus support the shift to subjective universality, the fact that all were present in varying degrees prior to the *Remarks* while Kant still advocated objective universality as

---

a moral standard suggests that external influences may have played at least some role in Kant's shift.

<sup>15</sup> Kant's personal library included a 1756 German translation of the 1751 edition of Hume's *Enquiry Concern the Principles of Morals* and German translations (from 1760 and 1762) of Hutcheson's *An Inquiry into the Original of Our Ideas of Beauty and Virtue* (1725) and *An Essay on the Nature and Conduct of the Passions, with Illustrations of the Moral Sense* (1728). See Warda 1922 and Kant 1992a, pp. 502-3.

<sup>16</sup> Hume 1975, p. 230.

<sup>17</sup> Hume 1978, p. 581, Hume's emphasis.

<sup>18</sup> Smith 1982, p. 46. This particular quotation is from an addition to the second edition, which Kant could not have read at this time, but the difference between Smith and Hume on sympathy is clear from the first edition, which was translated into French in 1764.

<sup>19</sup> In the *Enquiry*, Hume writes,

General language, . . . being formed for general use, must be molded on . . . general views, and must affix epithets of praise or blame, in conformity to sentiments, which arise from the general interests of the community . . . Sympathy . . . is . . . fainter than our concern for ourselves . . . but for this very reason, it is necessary for us. . . to neglect . . . differences and render our sentiments more public and social. (Hume 1975, p. 228)

Moreover, Manfred Kuehn has argued that Kant had access to some of the arguments of Hume's *Treatise* through his friend Joseph Green, a British merchant based in Königsberg (see Kuehn 2001, pp. 265, 272-3, 482n64).

<sup>20</sup> The translation was prepared by Marc-Antoine Eidous, Baron d'Holbac and published under the title *Métaphysique de l'âme* in 1764. The review appears in F-M de Grimm's *Correspondance litteraire* (Part 1, vol iv, 291f.; March 1, 1765), which was widely read (and of which Diderot later became an editor).

<sup>21</sup> Even if Kant *could* have read Smith, of course, there is no direct evidence that he *did* read Smith. Kant does not, of course, mention by name everyone that he reads, but he does mention Rousseau, Hutcheson, and Hume, and in his "Announcement," Kant strikingly lists "Shaftesbury, Hutcheson, and Hume" as having "penetrated furthest into the search for the fundamental principles of morality" (UD 2:312), thereby naming all major British sentimentalists *except* Smith. Of course, this "Announcement" may have been written before Kant's later *Remarks*, and the absence of Rousseau in this list suggests that Kant is not revealing all his most important influences from the mid-1760s.

<sup>22</sup> *Remarks* includes over 20 direct references to Rousseau or his works. See, e.g., Bem 20:9, 14, 17, 29-30, 42-4, 48, 50, and 58.

<sup>23</sup> See Rousseau 1979, pp. 461-2 and Rousseau 1997, pp. 52, 59.

<sup>24</sup> Rousseau 1979, p. 460, cf. Rousseau 1997, p. 50.

<sup>25</sup> Rousseau 1997, p. 62.

<sup>26</sup> Rousseau 1997, p. 67, cf. Rousseau 1979, 462.

<sup>27</sup> Rousseau 1997, p. 63.

<sup>28</sup> Rousseau 1997, p. 54; see too Rousseau 1979, p. 461).

<sup>29</sup> In this context, it is not surprising that Kant later comes around to explicitly political metaphors for morals, of which *Groundwork*'s "kingdom of ends" is the most famous (G 4: 433).

---

<sup>30</sup> For the sake of space, this paper has not discussed the connection between Kant's rationalism and subjective universality. For some important *Remarks* relevant to this connection, see Bem 20:97, 145-6, and especially 167.