Chapter 3: The Categorical Imperative and Universal Law

The reciprocity thesis gives us the mutual entailment of morality and freedom, as captured in a ‘harmonizing’ of effects and intent both in ourselves and amongst the collective actions of moral agents. As I discussed in Chapter 2, the reciprocity thesis is a constitutive principle of experience, as it allows us \textit{a priori} knowledge of a grounding principle. This chapter examines Kant’s Categorical Imperative as the law governing human action and its role in bridging the conception of self as a member of both the empirical world and an intelligible realm. I will present three interpretations of the first, and most commonly referenced ‘universal law’ formulation of the categorical imperative, as I work towards legitimizing Kant’s claim that it is functionally equivalent to the ‘kingdom of ends’ formulation and the ‘humanity’ formulation. To this end, I take up Alan Donagan’s discussion of the matter. This will illuminate the concept of reciprocity at work in Kant’s moral theory in considering a human agent’s participation in a universal, noumenal, legislating community. This \textit{a priori} concept of collective ‘humanity’ results in the derivation of the moral law as a \textit{synthetic a priori} proposition for each human agent and, thus, will be critical to understanding the political requirements of morality.

3.1 Understanding the ‘Universal Law’ Formulation

The categorical imperative is, for rational human agents, the principle of action that guarantees the instantiation of the concept of reciprocity. Freedom and morality are themselves reciprocal concepts in the meta-ethical sense, and the categorical imperative provides the manifestation of that concept in direct application.
It is both the most accessible, in terms of quantity of secondary literature, and the seemingly least accessible, as regards the content of that literature, aspect of Kantian moral theory. Because the *Groundwork* introduces, develops, and exalts the categorical imperative as the supreme principle of morality, it is taken as the keystone of Kantian thought. Perhaps because the formulation of the categorical imperative at 4:402 appears first, or because it is frequently referred to, it has been the source of analysis and criticism focused on the criterion of ‘universalizability’ that it is understood to require. This has led some commentators to discount Kant’s moral theory as being too restrictive in its requirements, or unable to fully comport with commonly accepted moral requirements, or altogether devoid of content.¹ Allen Wood argues that the universal law formulation of the categorical imperative “tells us no more than that our maxims ought to conform to whatever universal laws there are”² although he says that Kant does not provide adequate means of discovering what those laws might be. Wood’s conclusion, therefore, is that the first, and most common form of the categorical imperative provides no particular content.

But this formulation of the categorical imperative must be recognized as an *initial* formulation. The priority of its presentation signals the full range of considerations that Kant describes as necessary to a ‘universal law.’ In order to be universal, a maxim must be unconditioned, and in order to be a law, a maxim must be objective. If the categorical imperative required mere ‘universalization’ as the only

¹ For an example of some of these issues raised, see Jeffrie Murphy *Kant: The Philosophy of Right St.* Martin’s Press, NY 1970.
² Wood, Allen *Kant’s Ethical Thought* Cambridge: Cambridge University Press 1999, p.81
guide to moral action, there would hardly be a need for Kant to provide additional formulations and he would have had no need for the claim of equivalency between those formulations that many have claimed is patently false.3 While the subsequent formulations will be taken up a bit later, I wish to dispense first with the claim that the initial formulation of the categorical imperative yields nothing but an empty formalism. Instead, by examining some competing interpretations of the first formulation, it can be understood to generate the very set of duties Kant intended.

The categorical imperative to “act only in accordance with that maxim through which you can at the same time will that it become a universal law”4 has been read, erroneously, to entail each of the following:

i. Act in such a way that you can also will that everyone is permitted to act similarly. This interpretation, evident in the work of G.C. Field,5 picks up on Kant’s restriction that we not make ourselves an exception to the rules of moral conduct. So, if I generally see merit in truth-telling but, because I’ve fallen on hard times, I am considering borrowing money which I know I cannot repay in the foreseeable future, I have to ask myself ‘would it be allowable for anyone, if he falls on hard times, to willfully mislead someone by borrowing money based on a promise to repay that is not likely to be kept?’ In answering myself I would be considering the forward-looking consequences of such a policy. If everyone were allowed to break promises if and when his particular situation, based on his own assessment of it, seems to

---

3 The equivalency claim appears at 4:436.
4 Groundwork 4:402
5 Field states it as “what is sauce for the goose is sauce for the gander,” in “Kant’s First Moral Principle” in Mind, New Series, vol. 41, no. 161 Jan. 1932 pp.17-36.
warrant it, then the resulting condition would be unsustainable. Promise-making requires some level of trust and meeting of the minds with regard to the agreement. Allowing exceptions, especially at the discretion of the individual placed under obligation, would undermine the very conditions required for successful promise-making. I will refer to this particular reading of the categorical imperative as the ‘test for a self-defeator.’

ii. Act in such a way that you would be comfortable with others adopting such a maxim towards you. This interpretation offered by Sidgwick might appear to apply Kant’s key notion of reciprocity in moral conduct. However, it is instead only a slight variation on interpretation (i). Here, in addition to asking if willing that $x$ defeats your own purpose, it is being asked ‘are you okay with someone willing that $x$ in turn?’ So, returning to the example of borrowing money on false pretenses, the question is not just if you would frustrate your own effort to borrow money from another, but instead, would you tolerate someone borrowing money from you and not paying it back? Or, is it okay to will that you deny help to a drowning person regardless of the ease of the potential rescue, knowing that you would then be passed by if an off-duty lifeguard saw you drowning in a kiddie pool? The relevant consideration then becomes not about what it is right to do, but what actions should be avoided such that they don’t give license to others to do the same. The ‘golden rule’ is often misread in this purely self-interested way and Kant himself objects to equating the categorical imperative with that golden rule.\(^6\)

\(^6\) *Groundwork* n. 4:430
iii. Act only according to a rule that you can formulate as a ground of permissibility for action in a general sense. This reading of the categorical imperative, as I understand Nicholas Rescher’s quasi-consequentialist interpretation\(^7\) to support, assumes that an agent should be other-regarding in his decision making. In the case of borrowing money, the money’s intended use and harm to the lender would be relevant considerations. For instance, borrowing money to purchase necessary medicine for my ailing brother, with knowledge that repayment is unlikely, might be permissible where money to be wagered at the racetrack might not be. The first step in this moral analysis is a weighing of the benefit sought by the borrower. The second step under this interpretation of the categorical imperative is an assessment of the harm to the lender. In surreptitiously borrowing five dollars from an uncle who just cashed in the sole winning ticket for the Mega Millions jackpot, I am (under this view) perpetrating much less harm than if I were asking a widowed neighbor for the entirety of a given month’s social security check. In fact, it might be argued in this same vein that borrowing five dollars from the rich uncle to cover the copay on that necessary medicine for my ailing brother would do no harm at all if that particular five dollars had been earmarked by my uncle to be cashed in for quarters to be thrown off the top of the Empire State Building onto unsuspecting pedestrians below. The categorical imperative in these cases is thereby reinterpreted along the lines of ‘act only in such a way that others should do the same in like cases.’ According to Rescher,

\(^7\) Rescher, Nicholas *Kant and the Reach of Reason: Studies on Kant’s Theory of Rational Systematization* (Cambridge, Cambridge University Press, 1999), specifically chapters 8 & 9.
Kant’s emphasis on the rationale of action led him to move in a somewhat idiosyncratic direction and to shift from the moral acceptability of acts as such to the viability of the maxims (or precepts) that characterize modalities of action…the modality of action [involves] three component elements: an action or action type, a rationale (reason, motive, ground) for performing the action, and an injunction or instruction that is to be done as a general rule in appropriate circumstances.8

Rescher goes on to call Kant only a quasi-deontologist, and concludes that Kant’s moral theory takes the form of a ‘hypothetical consequentialism’ because “for him, duties themselves follow from the contemplation of consequences.”9 This interpretation takes the criterion of universalizability as a requirement to choose the best course of action in a particular set of circumstances. Should I will simply that I may borrow money without intent to repay so long as my intended use of that money is good or is it more complete and therefore better to will that I may borrow money without intent to repay so long as my intended use of the money is good and it causes no great harm to the person I am borrowing from? The latter is more complete as a rule, takes intuitively relevant considerations into account (under this reading) and can therefore be accepted as a guide to action. Rescher’s reading of the categorical imperative asks the agent to reflect on the situation and formulate a general principle according to which his intended action can be stated as a general permission to act.

My goal here is not to provide a refutation of these readings that is particular to each commentator, and I find Rescher, in particular, to offer substantive insight into Kant’s teleology and theory of moral motivation. Instead, by discussing them together, I wish to illuminate a feature shared by each that forms the basis of the

---

8 Rescher, p.213.
9 Rescher, p. 218.
contemporary consequentialist reading of Kant that builds upon the sentiments of John Stuart Mill, C.D. Broad, and R. M. Hare. The feature I speak of is an outcome-based approach to ethics. Each of the formulations given requires the agent to use imagined future states as a guide to action. The test for the self-defeator, which comes closest to the true intent of the categorical imperative, misses the boat by looking to what would be if an agent wills x, rather than analyzing what is. The outcome-based approach claims that to will that a promise be broken in a particular circumstance is self-defeating solely because it leads to a breakdown of trust and to the future impossibility of willing the same, either myself or by another. To accept this as the proper interpretation of Kantian morality is to reject a fundamental precept—that I conceive of myself as a lawgiver. If I accept the concept of self-legislation that Kant stresses as fundamental to his view, then a maxim of the promise-breaking sort is self-defeating in the moment willed—whether acted upon or not—because it creates a tension between my contingent volition and the necessity of my capacity for lawgiving which is inherently (and instantly appears as) self-contradictory. In other words, the promise-breaking maxim requires that I affirm promise-keeping as a universal objectively-valid maxim while simultaneously exempting myself from that principle’s reach as I seek an exception to the rule. I would be constructing a set which, in order to be complete, requires me as a member. But at the same time, I would be considering myself removed, contingently, from the set. Both cannot hold. It is therefore a logical impossibility that I am both a law-giver and a law-avoider in the same instance. And this captures Kant’s basis for
rejecting all outcome-based approaches to ethics in general. Not only is some future state of affairs a contingent consideration (as a result of time elapsed) but a forward-looking assessment is necessarily based on a weighing of individual reactions to possible future states which would make such a determination of action not only contingent but also subjective, and thereby not law-governed.

A closer examination of the categorical imperative’s purpose in Kantian ethics will assist in the derivation of particular duties from it. Kant is seeking a pure moral philosophy not specific to humans, but applicable to all rational beings and, as such, it must be found “a priori simply in concepts of pure reason.”

10 The categorical imperative’s appeal to us to formulate our maxims as universal laws is an appeal to formulate our maxim in accordance with the a priori concept of freedom. To do so requires that we abstract from all empirical, merely contingent considerations and instead focus only on what is contained in the complete concept of autonomy. This, for Kant, requires a consideration of rational beings as self-legislating members of a kingdom of ends. 11 This consideration is necessarily time-independent. An agent’s maxim cannot be considered with regard to a future-state for two reasons. First, Kant says that the very concept of obligation imparts to us an “idea altogether beyond the chain of nature” and time is the measure of such a chain. 12 This comports with his distinction between what is and what ought to be. If it were the case that efficient causation was sufficient causation of all things, then it would be impossible to

\[\text{\textsuperscript{10} Groundwork 4:389.}\]
\[\text{\textsuperscript{11} A more complete discussion of this can be found in the next section.}\]
\[\text{\textsuperscript{12} Kant. Review of Schultz, 8:13.}\]
conceive of things being otherwise. But as Kant says, if rational agents ought to act in particular ways, then they can do so, which is another way of affirming the productive power of reason, of freedom positively conceived. Second, moral evaluation consists in an assessment of “the disposition, let the result be what it may.”\(^{13}\) So, in both the formulation and the assessment of an agent’s maxim, time-dependent considerations are irrelevant. According to this, Kant concludes that consequences are incidental and therefore not the proper subject of morality.

A Kantian rejection of consequentialist misinterpretations of the categorical imperative is fundamentally based on such interpretations’ reliance on contingent circumstances and individual reactions to hypothetical future-states. That would create a morality based not on principles, but on prediction; not on necessary and universal concepts, but on contingent and subjective propositions. A moral agent can formulate maxims of action and maxims of ends, but he cannot formulate maxims of outcomes. Therefore the consequences of his actions, be they considered pleasant or unpleasant, are not the proper basis of moral judgment. Kant is seeking an a priori understanding of the moral law, and that requires, just as natural laws do, insight into the objective reality of a concept in conjunction with an intuition. The category of ‘community’ from the first Critique is at work here via the concept of humanity. It is through that concept that we will see the true equivalency of the three formulations of the categorical imperative.

\(^{13}\) *Groundwork* 4:416.
3.2 Questioning the Equivalence of Kant’s Three Formulations

The categorical imperative is given to us in the *Groundwork* via a number of formulations that Kant asserts are “at bottom only so many formulae of the very same law.”\textsuperscript{14} There is a great deal of secondary literature that seeks to explore the way in which the formulations might be seen as logically, or at least functionally, equivalent. That is, an explanation that the concepts entertained in each formulation are reducible to a logically equivalent form or that each formulation generates, in a unique way, the same set of duties. The conclusion of this analysis appears to be, as Alan Donagan puts it, that “Kant was simply mistaken”\textsuperscript{15} in his assertion. Donagan focuses his attention on the first and second formulations:

C.I.1. [Universal Law Formula] *Act only in accordance with that maxim through which you can at the same time will that it become a universal law.*\textsuperscript{16}

C.I.2. [Humanity Formula] *So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means.*\textsuperscript{17}

According to Donagan, the humanity formula generates an obligation to help others in the specific sense of making “some rational plan to do what you can for [those who need help] without neglecting yourself or others who have more pressing claims,” an obligation, he continues, that “does not follow at all from the universalization test.”\textsuperscript{18}

\textsuperscript{14} *Groundwork* 4:436.
\textsuperscript{15} Donagan, p. 152.
\textsuperscript{16} *Groundwork,* 4:402.
\textsuperscript{17} *Groundwork,* 4:421.
\textsuperscript{18} Donagan, p. 152.
Donagan describes the ‘universalization test’, in apparent contrast to the outcome-based approach I discussed previously, as asking “whether we could will that everybody in that sort of situation should act as we propose to.”\(^{19}\) But, if it can be accepted that morality is the doctrine of free action and the maxims of free action are aimed at preserving, if not also promoting, human freedom, then C.I.1 generates an obligation to help others in precisely the same way as C.I.2. This method of duty-derivation will be explained in further detail a bit later, but Donagan continues in his claim that the obligation would not be of the form ‘help others at all times’ because there may be instances when helping others hurts me.\(^{20}\) Therefore, according to Donagan, any maxim derived that does not take that form fails to be ‘universalizable’ in the way he asserts as relevant. But how this is so, given the moral law’s grounding, is not at all clear. Donagan has thereby incorrectly rejected C.I.1’s ability to generate appropriately other-regarding duties.

This brings me to Donagan’s second claim, that C.I.2 requires that we help any and all persons so long as our own needs and the ‘more pressing claims’ of others are not neglected. Kant certainly identified a duty to help others as fundamental to morality. However, the duty is not generated exclusively, nor as generally, from C.I.2 as Donagan argues. Because the moral law rests on an \textit{a priori} concept of freedom, C.I.1 also gives human agents the duty to protect and promote that very freedom for all rational beings to the fullest extent for the sake of our noumenal community, which consists not only of ourselves, but others as well. But Donagan is

---

\(^{19}\) Donagan, p. 145.
\(^{20}\) Donagan, p. 152.
speaking of people who need help in more specific ways such as a ride to the doctor’s office or a warm meal at Thanksgiving. It will be argued in the next chapter that we do have a duty to promote the happiness of others. But subjective, contingent considerations are not the hallmark of C.I.2. C.I.2 frames action in terms of what is owed to all human agents by virtue of their humanity, not what is owed to particular agents based on their individual projects. C.I.2 requires universality, objectivity, and consideration of the plurality of agents as C.I.1 does. If Kant was not mistaken in claiming the formulations of the categorical imperative to be at least functionally equivalent, and C.I.1 and C.I.2 necessitate the same actions (that is to say, against the tradition Donagan takes up), why did Kant see fit to complicate matters by offering multiple presentations of the categorical imperative? Those who reject Kant’s equivalency claim cite the various formulations as evidence of their difference, but looking at each formulation in turn offers an answer to that question and yields a more complete understanding of the nature and function of the moral law in Kant’s writing.

Having rejected the future-dependent interpretation of C.I.1, we know that universality, as opposed to universalizability, is the relevant criterion. The relevant consideration of a maxim not being the quantity of human agents who can will it, but rather the quality of the principle being willed. C.I.2 generates duties in a way that sheds additional light on the ‘universal law’ condition of the categorical imperative. The humanity formula concerns itself with the function of the moral law within rational beings whose wills are not perfect. As Kant describes, beings with a ‘holy
will’ act in direct accord with what they will. That is due to the wholly rational nature of such being, and with regard to God, to the productive power of pure reason. But humans are created and finite creatures who have our reason’s power tempered by the senses. Actions for humans are determined both by principle and inclination. We have the unique challenge (as opposed to divine beings) and unique ability (as opposed to non-rational animals) of discerning appearances from things in themselves. In considering our own character, we can consider ourselves as we appear in the world—subject to natural, efficient causation and therefore heteronomous—or we can, as C.I.1 requires, consider ourselves as a lawgiver beyond that control. Insofar as we recognize ourselves as capable of self-legislation and accordingly effective in the world, we are autonomous. The recognition of the self as self-legislating is necessary to generate the ‘oughts’ that guide our action. To preserve that capacity, Kant requires that we act in a way that respects the non-empirical part of ourselves. The essence of that non-empirical aspect of our being is the participation in the intelligible, non-contingent, objective realm. This is a realm of abstraction, of universal rationality. The ability to participate in this realm while still maintaining a presence in the empirical realm is the essence of our humanity. It is only the intelligible realm that grounds free action. It is only our intelligible character, as linked to a larger whole, that provides an objective value to our being. Therefore, that aspect of self is the source of worth in ourselves and in others. Recognizing the humanity in myself as an end in itself means that not only must I respect my own capacity for free action, but I must not allow another to disrespect it.
Likewise, in recognizing that others have this capacity, I realize that they have an objective value which I respect in the same way I respect myself. For what the humanity formula presents to us and demands from us is that we respect the absolute worth of rational, intelligible beings. To the extent that we must regard ourselves and others as having a rational, intelligible nature we must act and expect others to act with the respect due to that nature, treating it as an end in itself and not merely as a means. This provides a limiting condition on my actions such that I cannot treat humanity, in myself or others, as something that may be diminished in pursuit of another end. Humanity, as an end in itself, is of absolute worth, so there is no other end that would be ‘more worthy’ of pursuing at humanity’s expense. In this way, the categorical imperative requires that my maxims conform with the necessary and universal concept of humanity.

The third formulation of the categorical imperative appears at 4:438:

C.I.3. [Kingdom of Ends formula] Act as if [you] were by [your] maxims at all times a lawgiving member of the universal kingdom of ends.

The universal law formula of the categorical imperative makes it clear that our acts of will must not be based on contingent, individualized purposes, but instead on objective principles that could be held in common by all rational beings. The humanity formula makes known the absolute worth of rational beings as ends in themselves and the derivative requirement that we respect that aspect of ourselves and of other rational beings. The formulation of the categorical imperative as legislation within a kingdom of ends clarifies the scope of our free choice. We acknowledge,
through C.I.2, that humans participate in the intelligible realm. However, we are also aware (perhaps primarily so) of our participation in the world of sense. Because we are not wholly members of the noumenal realm, we have no claim of supremacy over any other person. Therefore, when we consider ourselves as members of a kingdom of ends we are commonly self-legislating, but as we are not sovereign over any other person, neither is any other person sovereign over us. To this point, it seems that Kant holds a radically individualized view, giving each person a right of self-determination consistent with respect for the self-determination of others. But it would be misreading Kant to ascribe to him a libertarian-esque view of human action. The libertarian doctrine assumes that every person should be free to do whatever he wishes so long as no direct harm comes to another. But Kant, by contrast, is again committed to a notion of human interaction, and is therefore very much influenced by the work of Rousseau. The ‘kingdom,’ for Kant, is “a systemic union of various rational beings through common laws”\footnote{Groundwork, 4:433.} much like Rousseau’s. And in that way, our membership in a kingdom of ends is a communal doctrine of freedom, through which our choices and others’ are reciprocally binding on all as a unified set of determining principles. The concept of a kingdom of ends thereby designates a unity of rational agents whereby all persons, by virtue of their humanity, are equally legislating members. This provides the universal scope of moral law as applicable to all persons by virtue of their supersensible nature.
When Kant states that each of the three formulae above is but a restatement of the same law, he does admit of a difference in purpose for each formulation. As we can see, C.I.1 provides the form of the moral law as universality (not a universalizability test in the commonly held sense); C.I.2 provides the matter of the moral law as an end in itself; C.I.3 demonstrates the totality of the moral law as determining all lawgiving. The concept that inheres in each formulation is the concept instantiated by the moral law: freedom. The categorical imperative, fully understood, is the prescription for free action. Thus, autonomy of the will, and the supreme principle of morality ascribed to it, is the culmination of the three criteria of freedom:

1. that we hold respect for humanity in ourselves and others;

2. that we consider our rational nature as united with the rational nature of all persons;

3. that we will only what conforms fully to 1 and 2 taken together.

The reciprocal nature of freedom and morality is therefore established as moral action requires self-legislation and true self-legislation requires the objective determinations of the moral law. Henry Allison is correct in his analysis of this ‘reciprocity thesis’ in citing its two significant consequences: first, that while the moral law is a synthetic a priori proposition, “it would be analytic if freedom of the will were presupposed” and second, that the reciprocity thesis “entails that freedom of the will is not only a

---

22 *Groundwork*, 4:436.
necessary but also a sufficient condition of the moral law.”23 Attention to Allison’s first observation is critical to deriving particular duties from the categorical imperative and also saves Kant’s theory from Sidgwick’s problem of the imputability to agents of immoral actions.24 The moral law is a synthetic a priori proposition for humans because while the idea of freedom allows us to think of ourselves as members of the supersensible, intelligible realm, we always retain an intuition of ourselves as members of the empirical, sensible world. So, my productive power of willing generates a categorical ‘ought’ which tells me what I can do, but only if I choose further to be motivated solely by that sense of duty and not by other inclinations. If there were a wholly intelligible being, a being with a purely rational will, then whatever that being willed would be done, necessarily. A purely rational will does not have any imperfections that lead to a conflict between the will and volition. This is Kant’s notion of a holy will, an attribute of God, and only in this state of moral perfection is the moral law an analytic a priori proposition. It is the synthetic character of the moral law in humans that generates and categorizes our duties, beginning with the duty of self-knowledge.

The formulation of the categorical imperative and Kant’s various presentations of it lays the groundwork (as the title of the work introducing those formulations promises to offer) of a content-ful moral theory. While Kant would later argue that political revolt is contrary to duty, it is easy to see why he embraced the theoretical underpinnings of the French Revolution. The Revolution’s motto—

23 Allison, p. 274.
24 Guyer, p.xlv.
‘Liberty, Equality, Fraternity’—parallels the nature of human obligation and the concept of autonomy. The universal law formula of the categorical imperative is concerned primarily with establishing the liberty of every agent by virtue of his rationality. The humanity formula establishes the equality of rational agents as ends in themselves and prohibits the consideration of one’s own ends at the expense of another’s. C.I.3 establishes each rational agent’s membership in a kingdom of ends, requiring consideration of the fraternity of agents. As promised, these three formulations are the instantiation of the meta-ethical concept of reciprocity. Against Leibniz, Kant has given us a community conceived by the understanding in conjunction with an outer intuition in space. Namely, each human agent via recognition of his own humanity can come to understand both his power and his obligations derived from the moral law. We now must turn to the application of these concepts in the realm of human interaction.