Chapter Nine
The Principles of Practical Reason

> The principle of morality is thus the idea of a will, insofar as it is a law to itself. (29:628)

a. One Principle, Many Guises (Revisited)

As discussed in chapter seven, according to Kant, the activity of any cognitive faculty – theoretical or practical – can be characterized in three interrelated ways. First, since all cognition involves a representation or representations in the subject, it can be described in terms of the subject's representations and attitudes and their (internal) relations to one another. Second, since all cognition has an object which it represents as something that either exists (in theoretical cognition) or ought to exist (in the practical cognition), it can be described in terms of its proper object(s) and their relations. And third, since cognition requires a certain sort of relationship between the subject's representations and their objects, any such faculty's aims can be described in terms of the relation between the subject and these objects that obtains when this faculty functions properly.

At the same time, by the end of chapter eight, we have arrived at least *five*, closely related ways of characterizing reason's fundamental end or function:

> (i) Reason's proper activity consists in (theoretical and practical) **mediate inference**.
> (ii) Reason's proper activity aims at (theoretical and practical) **cognition from principles**.
> (iii) Reason's proper activity aims at (theoretical and practical) **comprehension** (or, in the practical case, **practical wisdom in action**).
> (iv) Reason's proper activity consists in its **autonomous self-determination.**
> (v) Reason's proper object is the **highest good** (especially insofar as this unifies the three **ideas of reason** into a single systematic whole).

As we saw in chapter seven, to understand the systematic relationships between these characterizations of reason, it is helpful to organize them in terms of the tripartite division between three basic ways of thinking about any rational or cognitive capacity. In doing so, we can extend our treatment in chapter seven as follows:

> (i) We might characterize reason primarily in terms of the **rational subject** and their internal activities and representations – e.g., in terms of the idea of reason's **internal activities as autonomous** or its representations as standing in **inferential relations** to one another.

(ii) We might characterize reason primarily in terms of the **objects** of reason – e.g., in terms of the idea of theoretical reason as taking as its ideal object the three **ideas of reason** or of theoretical and practical reason *together* as taking as its ideal object the **highest good**, as a system that relates together our practical objects or ends and the various ideas of theoretical reason.

(iii) Finally, we might characterize reason primarily in terms of the proper **relationship** between the subject and these objects – e.g., in terms of the idea of reason as the capacity for **(theoretical and practical) comprehension** or (alternatively) for as the capacity for **autonomous action** (both internal and external).

In this chapter, much as in chapter seven, I plan to develop this interpretative framework by exploring how it can be used – not just to organize these characterizations of reason's activity or function – but also to systematize the various "principles" and "formulas" that Kant associates with practical reason – and, in particular, the various formulations Kant provides of the moral law.[410] As was true in the case of the principles of theoretical reason, my underlying assumption will be that the various principles of practical reason are very closely related. After all, as Kant says of the three main formulas of the moral law: "The above three ways of representing the principle of morality are at bottom only so many formulas of the very same law, and one of them of itself unites the other two in it." (4:436) But the precise sense in which this is true will only emerge over the course of our discussion.

b. The Formula Of Universal Law

To explore these ideas, let's begin with the first of these main headings – that is, with characterizations of reason's activity which focus on what that activity aims to accomplish *within* the rational subject with respect to that subject's *representations* (considered as such). As we've already noted, in the theoretical sphere, the primary expression of this aspect of reason is what Kant calls the "logical maxim" of theoretical reason:

... the proper principle of reason in general (in its logical use) is to find the unconditioned for conditioned cognitions of the understanding, with which its unity will be completed. (A307/B364)

[410] It is interesting to compare this project with (Merritt 2018)'s discussion of how to derive all of Kant's various *virtues* from a general notion of good cognitive functioning (as this is manifested in the practical case). Although there are real differences between us, a detailed comparison of these two discussions would bring out many important points of convergence. This, it seems to me, only strengthens the general case for the interpretation of the moral law I will be developing here.

Thus, the first question we need to consider is which practical principle corresponds to this "logical maxim" in the practical domain. It is natural to think that the what is often called the "hypothetical imperative" must be the principle that plays this role:[411]

> Whoever wills the end, also wills (insofar as reason has decisive influence on his actions) the means that are necessary to it that are in his control. (4:417)

After all, just as the "logical maxim" asks us to *ascend* higher in the sequence of theoretical cognitions by searching for the condition of the conditioned, the hypothetical imperative requires us to *descend* lower in the sequence of practical cognitions by willing whatever is a necessary means to our willed ends. In other words, while the "logical maxim" asks us to find the unconditioned which underlies the determinate, conditioned objects of theoretical cognition, the "hypothetical imperative" asks us to further determine our general ends through translating these ends into more determinate actions, which are themselves conditioned by these general ends. In this way, while the "logical maxim" expresses theoretical reason's characteristic activity of searching for ever more general principles in order to explain the particular objects it cognizes – the "hypothetical imperative" expresses practical reason's characteristic form of activity – namely, its effort to translate its general maxims or principles into determinate choices and (ultimately) actions in the world. In this way, we might think of the differences between the "logical maxim" and the "hypothetical imperative" as stemming from the basic difference between the relationship with its object that is characteristic of theoretical and practical cognition respectively.

This does point to an important connection between these two principles. But despite this, they ultimately have quite different statuses for Kant. After all, as Kant stresses, the "hypothetical imperative" is an analytic principle. Thus, the validity of this principle (or principles) must follow in some way from an analysis of the concepts involved it.[412] Just how and why this is true has been the subject of much debate.[413] But it is

---

411 There is considerable debate in the literature about whether we should speak of a single "hypothetical imperative" in this context or whether it might be better instead to speak of a variety of more specific "hypothetical imperatives". I set this issue aside here, since it is incidental to my aims. See, for example, (Kohl 2018) for further discussion.

412 Of course, even this claim is not entirely uncontroversial. For example, (Wood 2007) writes, "This is not a theoretical but a practical proposition. It is not the proposition that he who (really and truly) wills the end also does in fact will the means, but rather the practical proposition that in performing the rational act of setting the end, I have thereby subjected myself to the practical norm that I *ought* to will the necessary means." (66) Such passages appear frequently in the discussion of these issues, but (as we have already seen) they rest a failure to appreciate Kant's account of how imperatives or practical propositions are so much as possible – an account that rests on a close connection between imperatives and corresponding theoretical propositions about our rational faculties. Here compare Wood's odd claim that the "immediate consequences" of "a rational act, an act of freedom" are always "not causal, but normative". (67) Once again, a central feature of Kant's account of these issues is that such rational acts have *both* causal and normative "consequences".

relatively clear, I think, that the "logical maxim" (in its full form) cannot be reduced to an analytic truth in this fashion. In particular, while it *is* an analytic truth that anything conditioned must (as such) be conditioned by a further condition – and so that any conditioned cognition must have some further condition, this alone does not demand of us that we *search* for this further cognition in the manner the "logical maxim" demands of us. And, more importantly, even if this analytic truth does establish that every cognition must either be (i) conditioned – and so have some further condition – or (ii) "unconditioned" in the minimal sense of *not being conditioned*, this does not show that every cognition either has some further condition or is "unconditioned" in the sense that is required in order to "complete reason's unity". After, as we have seen, in order to "complete reason's unity" we must arrive at a body of cognitions that form a genuine systematic unity – a body of cognitions which must all be grounded in a *unified system of principles*. And this involves considerably more than merely finding the condition for any conditioned. For example, by forming a disconnected set of cognitions, none of which one regarded as conditioned by any further cognition, one could in principle satisfy the demand to find the condition for any conditioned cognition. But one would not thereby have "found the unconditioned" in the sense required in order to "complete reason's unity". And so one would not thereby have complied fully with the "logical maxim".[414]

As this indicates, if the "logical maxim" merely demanded that we find the condition for any cognition which we regarded as conditioned, it would be quite similar in character to the hypothetical imperative. But, when read correctly, the "logical maxim" demands more than this – namely, it demands that we seek the *particular sort* of systematic unity which is required if reason is to satisfy its aims. Thus, in commanding us to "ascend" upwards in the sequence of conditions in the *particular* manner it does, the "logical maxim" effectively commands us to continue this inferential activity until all of the cognitions of the understanding have been placed within a *single, unified system* – one which terminates in a unified system of absolute theoretical principles. In this way, the "logical maxim" expresses – not just the *understanding's* demand for logical consistency – but also *reason's* demand for systematic unity as well. And in doing so, what the "logical maxim" demands in the theoretical domain plainly goes beyond what the hypothetical imperative demands in the practical. For the hypothetical imperative, on its own, in no way demands of us that all of our practical cognitions form a *unified system* in anything like this sense. Rather, the hypothetical imperative merely demands that we be *locally* consistent insofar as some choice of means is already implicit in our choice of ends, plus our background theoretical knowledge.

---

413 See, for example, (Korsgaard 2008; Wood 2007; Kohl 2018; Hill, Jr 2012). For more discussion of the role of the hypothetical imperative in this context, see (Tenenbaum 2019; 2021).

414 Of course, it is also true that all of our empirical cognitions must as such, be conditioned by other empirical cognitions, according to Kant. But this itself is not an analytic truth, but rather one of the most basic synthetic *a priori* truths about our form of cognition. So appealing to this fact in this context will not help us show that the "logical maxim" is an analytic truth in the sense at issue here.

In other words, the hypothetical imperative does not demand that we achieve anything like full-fledged practical cognition from absolute principles. For, that of course is a role that Kant reserves for the moral law. But the "logical maxim" does *in a sense* express this demand in the theoretical domain – albeit only in a "logical" guise, which focuses on our representations as such, as opposed to how these representations in relate to their objects. Thus, if we are searching for a practical principle which corresponds to the *full* "logical maxim", we need to find a principle that expresses *practical reason's* demand for *systematic unity*, but does so in a "logical fashion" – that is, in a manner that focuses on the systematic unity of the subject's *practical representations* as such. The practical principle which best meets this description is the formulation of the moral law that Kant calls the Formula of Universal Law (FUL):

> Act only in accordance with that maxim through which you at the same time can will that it become a universal law. (4:421)

So how exactly does the FUL fit into the present account? In the last chapter, we saw that practical reason aims at a certain sort of autonomy, and therefore at a form of "practical comprehension", "practical wisdom", or "practical cognition from absolute principles". Thus, a helpful way to think about the FUL is to see it as specifying a basic requirement on this particular sort of practical cognition.[415]

To understand why the FUL has this status, let's briefly run through what is required if a maxim is to count as an instance of practical cognition from absolute principles (in the sense we have explored above). First, a maxim of the will can only have this status, if it is derived from some general principle. And, second, it must ultimately derived from a principle which has that status, not merely "comparatively" or relative to some other cognitions, but also "absolutely". That is, the maxim of my will can only count as an instance of practical cognition from absolute principles if it is derived from a principle whose validity is not conditional on any further, more fundamental principle. Thus, for a practical principle to count as an absolute principle, it must be valid of practical reason in general without any restriction or further condition. And *this* will only be true if *this absolute practical principle* passes the FUL test as laid out by Kant. For, contrary to (say) self-effacing forms of consequentialism, Kant's conception of reason as essentially self-conscious commits him to the claim that any being with the capacity of practical reason is capable of becoming conscious of the principles that characterize practical reason as such. Thus, if X is an absolute practical principle in Kant's sense, which holds unconditionally for every rational will, any being with practical reason must be capable of being

[415] See (Engstrom 2009) for a deeply insightful discussion of these issues, to which I am much indebted, even though I develop these ideas differently than he does there. Also see (O'Neill 1989; Korsgaard 2009), although neither focuses on "cognition" in this context. Compare the discussion of (Rosenkoetter 2011), although I would not put things in terms of "reference" in the manner Rosenkoetter does.

consciousness of this fact. And to be *practically* conscious of the universal validity of this principle for practical reason just is to treat X as determining (at least in part) what any rational being ought to do. Thus, to have a practical consciousness of X's universal validity, just is to will that X holds with full universality. In short, for Kant, if some practical principle holds absolutely, any being with practical reason must be able to be conscious of it as having this status simply in their possession of this capacity. And if this is true, then any practical reasoner must be capable of willing that this principle governs the will of every practically rational being.

Thus, the maxim of our will can be an instance of practical cognition from absolute principles only if it is at least *based* on an absolute principle – and any such absolute principle must *itself* pass the FUL test. This, I think, should be *relatively* uncontroversial. The difficult question is whether being *derived* from such an absolute principle is sufficient to entail that the "maxim of my will" *itself* passes the FUL test in the manner the FUL demands.[416] The secondary literature with respect to this question is, of course, vast. But much of this literature proceeds in what seems to me a rather odd fashion. For, it is reasonably clear that Kant takes the FUL test to follow quite quickly from the idea of practical cognition from absolute principles.[417] And it is also clear that much of the terminology in the FUL – such as the crucial notion of my "maxim" in acting – is at least somewhat technical and so subject to interpretation. As a result, it seems to me that our primary goal in interpreting Kant's reference to (say) the "maxim of my will" should be to interpret this terminology in a manner that vindicates Kant's belief that the FUL test expresses an immediate consequence of the idea of practical reason as the faculty for practical cognition from absolute principles. Insofar as an interpretation of the FUL fails to accomplish this task, it seems to me that it should be rejected.

Moreover, it seems to me that it is not terribly difficult to provide an interpretation of the FUL that delivers this result. For we can do so simply by understanding Kant's references to the "maxim of one's will" in the FUL to refer – not to any old "intention" in the contemporary sense of this term – but rather to the whole system of maxims and principles that form one's full (implicit or explicit) grasp of the reasons why one is acting as one does. When read in this way, the "maxim of one's will" will often (implicitly) be extremely complex. But Kant never claims otherwise. And when we read the FUL in this fashion, it turns out that Kant is entirely correct to regard the FUL as a straightforward consequence of his conception of practical cognition from absolute principles. After all, when interpreted in this way, the "maxim of my will" will indeed pass this test just in case it is *properly derived from an absolute principle*. For, as we have just seen, any such absolute principle must pass the FUL test – and so too must any maxim which is properly derived from such a

[416] Compare (Hill, Jr 2012; Wood 1999). For some recent discussion of this issue see (Timmons 2008; Nyholm 2015; Kleingeld 2017). As we will see, my reading is in some ways closest to Kleingeld's.
[417] (Engstrom 2009; Reath 2010; 2013b).

principle *insofar* as we understand this maxim to itself encode within it the manner in which it derived from the absolute principle in question – and so, any conditions on my action that follow from this.

As we have just seen, one advantage of this interpretation is that it makes it easy to understand Kant's insistence that there is a very close connection between the FUL and the idea of practical cognition (or willing) from absolute principles. But this interpretation also deals with many of the most familiar alleged counter-examples to the FUL.[418] For instance, it seems obvious there are many perfectly permissive "local" maxims of action, which it would be insane to view as universal laws that apply to every rational being simply as such. So if the FUL demands of me that I treat these "local maxims" as universal principles of rationality, it seems hopeless as a principle of practical rationality.

Fortunately, for Kant, this objection (and others like it) rests on a failure to understand exactly what, for Kant, is involved in accepting something as one's maxim for action. For example, suppose I intend to perform some action that is harmless if only I perform it, but harmful if everyone does so. Then, if I am rational, my intention to perform this action must be implicitly conditional. For example, in this case, if I am rational, I will only intend to perform this action on the condition that everyone else is not intending to do the same. But then, on the reading of "the maxim of my will" just laid out, this condition must also be part of the *full* maxim upon which I am acting if I am rational. And it is this complete "maxim of my will" that we must consider when applying the FUL. But when we do so, we quickly see that this (conditional) maxim is harmless when universalized. Thus, when read in this way, this species of counter-example to FUL, at least, disappears.

Of course, this does not respond to *every* objection to the FUL. In particular it does nothing to respond to the worry that the FUL is empty – or, more modestly, that it (at least) fails to *rule out* many morally impermissible courses of action. But this is a much less fundamental objection to the FUL that the prior one. For it is an objection – not to the validity of the FUL – but rather to the claim that *all* of morality can be derived from it alone. And, in considering this worry, it is important to stress that Kant never claims that the process of moving from the FUL to a full understanding of morality's demands for any particular agent is at all simple or straightforward. Rather, Kant's position is that the FUL is indeed highly indeterminate in its demands – as we would expect any universal principle of practical reason to be – and that it must, therefore, be made more determinate through the activities of practical reason if it is to play the role we need it to in our everyday moral life. So to point out the indeterminacy of the FUL (considered on its own) is simply to note an explicit

[418] See (Wood 1999) for an excellent overview of such potential counterexamples. And compare (Baron 1999) here. For the importance of the proper reading of "maxims" in this context, see (O'Neill 1989; H. Allison 2011; Nyholm 2015; Sensen 2014)

aspect of Kant's views. And so, in order to examine Kant's response to this worry, we would need to look beyond the topics under discussion here to consider in detail Kant's substantive normative ethics and political philosophy. Thus, for the "empty formalism" objection to really be an objection to the role of the FUL in Kant's philosophy, it would have to be shown – not that the FUL *alone* does not deliver all the moral verdicts we expect – but rather that the FUL *together* with the manner in which it is made determinate in our actual empirical-social-political context – fails to deliver these verdicts. And that is simply not an issue that can be settled by considering the FUL in isolation from (say) the manner in which it interacts with the nature of human psychology or the political conventions we as a society have formed in order to make its demands more determinate.[419]

In any case, my aim in this section is not to offer anything like a full defense of the FUL. Rather, my goal is much more modest – to explain why the FUL deserves to placed under the first of our three headings as something like the practical analogue of the "logical maxim". To recognize the plausibility of this claim, it suffices to argue that the FUL expresses a (i) necessary condition on practical cognition from absolute principles, which (ii) emphasizes the manner in which this requires systematicity in my practical cognitions, but (iii) does so by focusing on the relationships between practical representations or maxims (as such). I hope at least to have said enough to make each of these claims plausible.

Before moving on, I would like to compare my account with Guyer's recent attempt to explain how Kant derives the moral law from the nature of reason or rationality.[420] Much as is true of my account, Guyer's account focuses on reason as a faculty that consistently seeks sufficient reasons or conditions for conditioned cognitions so as to arrive a fully systematic body of both theoretical and practical cognition. Thus, for Guyer, reason is characterized by three main principles: the principle of non-contradiction, the Principle of Sufficient Reason, and a principle of systematicity (which Guyer sees as distinction from the second principle). In this, despite some differences of detail, Guyer and I are obviously in general agreement. But Guyer's understanding of the relationship between these principles and the moral law is nonetheless quite different from my own. In particular, Guyer believes that only the first of these aspects of reason – the Principle of Non-Contradiction – is directly relevant to Kant's derivation of the moral law. Thus, Guyer writes:

> Kant's derivation of the foundational principle of morality proceeds by the application of the principle of non-contradiction to the fact that every human being has a will of his or her own, in the form of the idea that to act immorally is both to assert and deny that the object of such an action, whether oneself or another, has its own will. This is a fact, in Kant's own terminology a "fact of reason", but it is not a

[419] For more on the political dimension of this, see (Pallikkathayil 2010).
[420] (Guyer 2019).

> mysterious moral fact, or a blue that somehow exists in the universe independently of our act of valuing it. It is simply a fact that cannot be denied on pain of self-contradiction... (Guyer 2019, 64)

Obviously, there's much in what Guyer says here that I agree with. But if Guyer means to claim that we can derive the validity of the moral law *solely* from the principle of non-contradiction, together with certain theoretical and non-moral claims about my nature and the nature of those around me, I must confess that this does not seem to me to be true to Kant's understanding of these issues.

Indeed, it seems to me that Kant would deny that any *practical* principle follows logically from these theoretical claims considered merely on their own. Rather, to move from a theoretical conception of myself and others to a practical conception of how I *should* treat such beings requires some appeal to a principle of practical reason – that is, to a principle that governs reason in its practical use. And while I agree with Guyer that the principle of non-contradiction applies equally to theoretical and practical reason, it does not seem to me that this principle, on its own, is capable of bridging this gap in Guyer's argument.

For these reasons, it seems to me much more plausible to follow Kant in viewing the moral law – not merely as following from with the principle of non-contradiction – but rather as the practical analogue of the Principle of Sufficient Reason. We've already seen why this is natural, given Kant's presentation of the relationship between reason and the moral law. And this connection will only become clearer in the next section, where we discuss the relationship between the moral law and the idea of humanity as an unconditioned end. Given that Guyer acknowledges that the principle of sufficient reason does play a positive role to play with respect to the Kant's practical philosophy, it is surprising to me that he does not draw these connections.

More precisely, as we have seen in our discussion of the "supreme principle" of theoretical reason, Kant understands the Principle of Sufficient Reason (in its legitimate form) to be an expression of reason's basic derive for cognition from absolute principles or comprehension. Of course, as we have already discussed, this drive is inevitably frustrated in the theoretical sphere by the limits on our capacities for theoretical cognition. But, as we discussed in the last chapter, the same is not true in the practical sphere. Rather, in this case, practical cognition from absolute principles *is* possible for us, at least to a considerable degree. And, insofar as this is possible for us, it is also possible for us to comply with the Principle of Sufficient Reason when this is conceived of as a principle governing *practical* cognition. Thus, we can read Kant's account of the moral law in general, and the FUL in particular, as an account of how it is possible for us to comply with the Principle of Sufficient Reason within the practical domain – even though such compliance is impossible for us in the theoretical sphere.

c. The Formula of Humanity

With that in mind, let's turn to the practical principles which focus on the "objects" of practical reason – that is, those that focus on the *ends* which practical reason sets for action. Here, of course, the fundamental principle for Kant is the Formula of Humanity (FH):

> So act that you use humanity, as much in your own person as in the person of every other, always at the same time as an end and never merely as a means. (4:429)

Much as was true of the FUL, I believe that this principle is best understood as specifying a requirement on *practical cognition from absolute principles*. But, whereas the FUL focused on how the idea of practical cognition from absolute principles constrains the relations between our representations or maxims, the FH focuses on the manner in which this idea constrains the proper objects of these representations – that is, the *ends* which are the objects of a body of practical cognition from absolute principles.

To understand how the FH does this, it will be helpful to return to the discussion of chapter two and, in particular, the idea that every rational capacity must (in a sense) treat its own proper activity as its most basic "formal" end – that is, as an end that gives teleological structure to all of the various activities of that capacity.[421] Given this general picture of rational capacities, practical reason must also treat its own proper activity in this way. And so, every principle of practical reason must be compatible with this formal end. So, a maxim of the will can only be a candidate for the status of practical cognition from absolute principles if it represents one possible way of making this "formal end" determinate.

In this way, every rationally permissible maxim is, for Kant, a way of making more determinate the abstract idea of practical reason as treating its own proper activity as an end in itself. This allows us to extract one basic end from the bare idea of pure practical reason. But the FH, of course, demands something considerably more robust than this. For the FH demands of me, not just that my faculty of reason treat *its own* activity in this way, but also that *everyone's* faculty of reason treat *everyone's* rational activities in the same manner. The gap between this two claims has been the focus of much of the critical literature concerning Kant's derivation of the FH. But the key to responding to it is already present in our prior discussion of the FUL. For, as we noted there, what we will can only be an instance of genuine practical cognition from absolute principles insofar as it is based on principles which apply to all practical reasoners without any

[421] (Engstrom 2009; Reath 2010; 2013b)

restriction or condition. And a principle can only have this status if the ends it invokes similarly hold for all practical reasoners unconditionally. That is, as Kant writes, "... if all worth were conditional and therefore contingent, then no supreme practical principle for reason could be found anywhere." (4:428) But if *any* end is unconditional in this sense, it will be the formal end that is characteristic of practical reason as such – namely, the end of respect for its own characteristic form of activity. Thus, if there are to be *any* absolute practical principles at all, the end of respecting practical reason in all its forms must hold unconditionally for all practical reasoners.

This point can be presented as an argument from the FUL to the FH, once we recognize that implicit in the FUL (as the principle of a rational capacity) is the idea that that capacity (like any such capacity) must treat its own proper exercise as its end. In saying this it is important to stress that I am not claiming that the FH is implicit in the FUL when it is read as an abstract principle in isolation from Kant's account of the nature of rational capacities. Rather, the FH can only be derived from the FUL when the FUL is understood *as the principle of a rational capacity*. In other words, to appreciate the connection that Kant establishes between the FUL and the FH, we need to consider what a rational capacity would have to be like in order to have the FUL as its principle. Thus, while there is a sense in which we can regard the FUL and the FH as "equivalent" to one another – on my account, this equivalence runs *through* the idea that both these principles are to be understood as expressions of the principle characteristic of practical reason *as a rational capacity*. In this way, the relationship between these principles only becomes clear if we consider them with the context of Kant's "capacities-first" approach to these questions.

Let's see how this argument from the FUL to FH via Kant's general account of rational capacities could be spelled out. We begin, again, with the idea every such capacity aims (as its formal end) at its actualization through its characteristic mode of activity. It is this, I want to suggest, that licenses Kant's claim that every rational being, simply in virtue of being rational, must treat *its own* rational nature as an end in itself. More precisely, as Kant makes clear, when he speaks of rational nature as an end in itself, he does not mean to characterize our rationality as an "end to be effected" (*zu bewirkenden Zweck*) – that is, as something that one aims to produce by one's action as a means to an end. Rather, our rational nature is an end in itself in the sense of being what Kant calls an "independent end" (*selbstständiger Zweck*) – that is, something for the sake of which one acts, even though it is not produced by one's action in this way. Thus, the crucial idea here is not that we should be engaged in the practice of promoting the existence of rational beings as such, but rather that we should act and reason in ways that manifest a respect for our own practical rationality. And the core of this respect lies precisely in the fact that practical reason (like any rational capacity) must represent its own proper exercise as an end to be respected in this sense. Thus, for Kant, every human being must necessarily

represent their own rational nature as an end in itself because this is simply part of what it is to be a creature who possess the capacity of practical reason.

In this way, by considering the nature of practical reason, we can extract from Kant's general account of our rational capacities the claim that, whatever else it wills, practical reason is always (at least implicitly) willing that it actualize itself via its own characteristic form of activity – or, alternatively, that practical reason must treat its actualization in this way as an independent end:

> *Own Rationality*: Practical reason must always treat its own proper exercise as an independent end. In doing so, whatever else it is willing at any given time, practical reason should be seen as (implicitly) willing that it actualize itself in its own characteristic mode of activity.

Once again, in reading Own Rationality, it is important to stress that it involves the idea of the proper exercise of practical reason as what Kant calls an "independent end" (*selbstständiger Zweck*) as opposed to as an "end to be effected" (*zu bewirkenden Zweck*). Thus, the idea that practical reason must treat its own proper exercise in this way does not imply, at least in the first instance, that practical rationality must be somehow focused on promoting the existence of reason in one's self or others.[422] Rather, what is at issue here is primarily the idea that practical reason (like any rational capacity) is always striving to act in a manner that is in accordance with its own distinctive form of activity.[423]

As we have seen, Own Rationality in this sense is a consequence of Kant's general conception of the nature of rational capacities. But this, of course, leaves us well short of the FH. Kant makes the transition from Own Rationality to the FH in a crucial, but highly compressed, passage of the *Groundwork*:

> The ground of this principle [the categorical imperative] is: *rational nature exists as an end in itself*. The human being necessarily represents his own existence in this way; so far it is thus a *subjective* principle of human actions. But every other rational being also represents his existence in this way consequent on just the same rational ground that also holds for me; thus it is at the same time an *objective* principle from which, as a supreme practical ground, it must be possible to derive all laws of the will. The practical imperative will therefore be the following: *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*. (4:429)

---

[422] For concerns about this reading, see (Wood 2007; Parfit 2011). Of course, Kant does take the FH to have implications for these questions as well, but these implications are downstream from the very basic end that Own Rationality attributes to practical reason as such.
[423] Compare (Darwall 2006).

So, how does the argument we are developing here fit with this passage? The argument we have just given for Own Rationality *at least* establishes that Own Rationality is a "subjective principle" for any finite rational being. For it certainly shows that any being with a capacity of practical reason will need to represent its own nature as rational as an end in itself in the sense just explained. But, in fact our argument thus far seems to show more than this. For, given this argument, Own Rationality's validity is grounded in features of practical reason that are shared among all practically rational beings as such. And so, the demand to "respect" our own capacity for practical rationality arises out of grounds that are "objective" in the sense of being rooted in the principles of practical reason itself – grounds that, therefore, characterize the nature of the proper object of practical reason, as opposed to merely characterizing the proper object of this or that individual's practical reason. So, there is a at least a sense of "objective" in which our argument establishes Own Rationality as an "objective principle" of practical reason as well.

But this still leaves us short of the FH – since Own Rationality says nothing about how I should regard the rational nature of other agents around me. Thus, we are left with the question of why Kant takes the FH to follow directly from the idea that something like Own Rationality is "at the same time an *objective* principle"? Or, perhaps better, with the question of what the notion of "objective" is which makes sense of this transition? This is one of the most debated questions in all of Kant's practical philosophy.[424] So I won't attempt to consider all the answers to it in the literature. Instead, I want to simply describe *one* way in which we might move from Own Rationality to the FH, making use of the fact that we have (at this stage of Kant's argument) already established the FUL as a principle of practical reason. I think this proposal is a reasonable reading of the text. And I think it is clearly available to Kant, given his other commitments. But I don't want to claim here that it is the *only* plausible way of reading this passage. For there is little doubt that text underdetermines the correct interpretation of Kant on this point.

So, on the interpretation I have in mind, how does the FH follow from the idea of Own Rationality as an objective principle of practical reason? To consider this, I want to consider how a principle like Own Rationality affects our understanding of the FUL as a principle of practical reason. Given Own Rationality, no matter what else I am willing at a given moment, I ought always to also be willing that I treat my own current rationality as an (independent) end in itself. So, this end will always be implicit in the full "maxim of my will", once this is completely spelled out. As a result, whenever I test a maxim M using the FUL test, what I really need to test for universalizability is ultimately not just M alone, but rather the package of M plus the

[424] See, for example, (Korsgaard 1996a; O'Neill 1989; Herman 1996; Wood 1999; Engstrom 2009).

maxim that I treat my own current rationality as an end in itself. In other words, once we consider it in the context of Own Rationality, the FUL test becomes the following:

> *Expanded FUL*: Act only in accordance with that maxim through which you can at the same time will that it *and* the maxim that one treat one's current rationality as an end in itself *both* become universal laws.[425]

But now, notice what happens when we move from the FUL to the Expanded FUL. Suppose, for example, that I want to see whether some maxim M of mine passes the Expanded FUL test. Then I will need to consider whether it is possible for me to consistently will M while willing that it *and* the maxim that one treat one's current rationality as an end in itself both become universal laws. That is, given our account above, I must consider whether it is possible for me to consistently will M while willing that it *and* the maxim that *one's practical reason actualize itself in its characteristic mode of activity* both become universal laws. Or, more simply, I must consider whether it is possible for me to consistently will M while willing that it *and* the maxim that *one act in accordance with the requirements of practical reason* both become universal laws.

But when is *this* something that I can consistently will? Well, if some maxim X is to be a universal law, it must at the very least be the case that every rational will conforms to X. Thus, a maxim M will pass this test only if everyone can (i) will M, and (ii) will that M is a universal law, and also (iii) treat their own rationality as an end in itself. But if everyone can do this, then it must also be true that everyone can simultaneously will *that I will M* while treating *their own* rationality as an end in itself. Or, alternatively, it must also be true of every rational being that they can (i) will that I will M and (ii) fully exercise their capacity for practical rationality in the proper fashion. For this is just a particular case of these more general claims. Thus, if my maxim M passes the Expanded FUL test, it will only do so because every other rational being can "rationally accept" that I adopt this maxim in this manner. Thus, the following principle follows from the Expanded FUL:

> *Universal Rational Acceptance*: Act only in accordance with a maxim which every rational being could will that I act upon while fully exercising their capacity for practical rationality.

If my actions are in accordance with Universal Rational Acceptance, then every rational being will always be capable of willing that I act as I do without in anyway deviating from the demands that practical reason places upon their own will. In this way, as we will discuss further in the next section, Universal Rational Acceptance

[425] This principle is, in many ways, similar to the reading of the FUL that has recently been defended by (Kleingeld 2017).

is very close to the ideal of "universal legislation" that Kant associates with the third formula of the moral law – and in particular, with the variant of that formula known as Formula of the Realm of Ends:

> FRE: Act in accordance with maxims of a universally legislative member for a merely possible kingdom of ends. (4:439)

But for present purposes, the important point is that Universal Rational Acceptance also provides us with a basis for deriving at least the core of Kant's understanding of the FH. For while there is of course much debate about the meaning of the FH, Kant often seems to regard the core of the FH to involve something very close to Universal Rational Acceptance.[426] For example, in discussing how to derive a prohibition of false promises from the FH in the *Groundwork*, Kant writes that,

> ... he who has it in mind to make a false promise to others sees at once that he wants to make use of another human being *merely as a means*, without the other at the same time containing in himself the end. For, he whom I want to use for my purposes by such a promise cannot possibly agree to my way of behaving toward him, and himself contain the end of this action. (4:429-430)

And in a related discussion in the second *Critique*, he writes,

> Just because of this every will, even every person's own will directed at himself, is restricted to the condition of agreement with the *autonomy* of the rational being, that is to say, such a being is not to be subjected to any purpose that is not possible in accordance with a law that could arise from the affected subject himself." (5:87)

As such passages indicate, Kant often seems to think of the essence of the FH (and, by extension, the FA) as lying in the idea that one should adopt a maxim only if every other rational being could will that one do so, while properly exercising their own capacity for practical rationality. And we have just seen that this principle follows from the FUL, once it is interpreted in the context of Kant's background understanding of the nature of rational capacities and principles. Thus, given that background, we have seen how to derive the FH (on one plausible interpretation of it) from the FUL, when the latter principle is interpreted as the principle of a rational capacity.

---

[426] Compare (O'Neill 1989; Engstrom 2009; Sensen 2011). For a contrary view, see (Wood 1999; 2007).

Given this, a helpful way to think about this transition from Own Rationality to the FH is in terms of reason's fundamental demand that we search for the further conditions or grounds that are required in order to complete reason's project of systematic unity. As we saw above, the FUL already expresses this demand – but does so merely in a "formal" fashion – that is, in a fashion that is concerned primarily with the relationships between our various practical representations considered as such. But once the FUL is paired with Own Rationality's claims about the necessary ends of practical reason, this demand for systematic unity takes on a "material" form as well. That is, it becomes a demand to arrive at a system of rational ends which is capable of satisfying reason's demands for systematic unity.

As this should indicate, Korsgaard and Wood, in their famous readings of the argument for the FH as a "regress of conditions", are right to see a close connection between the possibility of the moral law as an absolute practical principle and the possibility of unconditioned value.[427] Where readings like theirs go wrong is in imaging *this* relationship to be something we can discover via a *regress* on conditions. For to think of practical reason in this way is, in effect, to model practical reason's operations on the operations of theoretical reason – which does, of course, regress on conditions in just this way. But while this sort of "regress" is characteristic of theoretical reason in its canonical form, the canonical activities of practical reason move in precisely the opposite direction. In other words, unlike theoretical reason, practical reason has no need to "regress on conditions" in *search* of its fundamental principle – rather, it's reasoning *begins* with a basic consciousness of this principle (as Kant stresses in his discussion of the Fact of Reason). Thus, while the FH does connect the possibility of practical cognition from absolute principles closely together with the existence of unconditional ends, we should not see this connection as the product of the sort of *regress* of conditions that Korsgaard and Wood describe.

e. The Formulas of Autonomy and the Kingdom of Ends

In any case, with these connections between the FUL and the FH in mind, let's turn to the third of our major categories – namely, expressions of practical reason's principle that focus on the *relationship between both subject and object*. As we have discussed above, these will be principles that expresses reason's demand for systematic unity, not just as a demand on the subject's representations or on their proper objects, but also as concerning the *proper relations between* the subject and these objects. In this way, it is third category which presents practical reason's principle in its most complete form – for it unites together the ideas at work in the other two formulas in fully systematic fashion.

[427] See (Korsgaard 1996a; Wood 1999; 2007). For some criticism, see (Timmermann 2006; Hills 2005).

The main expression of this third perspective on practical reason's principle is to be found in the third main formula of the moral law, which Kant expresses (first) as the Formula of Autonomy (FA) and (second) as the Formula of the Realm of Ends (FRE):

> FA: Not to choose otherwise than so that the maxims of one's choice are at the same time comprehended with it in the same volition as universal law. (4:440)

> FRE: Act in accordance with maxims of a universally legislative member for a merely possible kingdom of ends. (4:439)

These principles are naturally grouped under this third heading, because, as Kant famously claims, they unite the first two formulas in a single principle by foregrounding what Kant, rather mysteriously, calls the "totality of the system" which combines together the "unity" emphasized by the first principle and the "plurality of ends" emphasized by the second.[428] In doing so, this third formula of the moral law brings to the forefront something that was merely implicit in both the first and second formula – namely, the moral law as characterizing a *unified system of relations between rational beings* considered *both* as practical subjects or agents *and* as practical objects or ends to be respected.

In doing so, it corresponds exactly to the third of our categories. In particular, as we have just discussed with respect to the FH, rational beings are (for Kant) both subjects who engage in practical thought and objects or ends that must be respected by such thought. Thus, the rational beings who are related to one another in the FA and FRE can be viewed either as subjects who are subject to the moral law (in the manner the FUL emphasizes) or as ends in themselves for that law (in the manner the FH emphasizes). In characterizing morality in terms of the proper system of *these* relations, the FA and the FRE are thus characterizing the moral law in terms of its implications for *the proper relationship between practical subject and practical object*. In this way, the FA can be thought of as expressing most completely the fundamentally intersubjective character of the moral law for Kant. For while all three formulas express this in some way, only the FA (and especially the FRE) does so in a manner that does equal justice to both of the two roles (rational subject and rational object, as it were) that rational beings play in Kant's account.

This is obscured to some degree by the fact that the same beings serve as both the subjects and the objects within these relations. But the basic point remains. For, by presenting the moral law in terms of the proper system of relationships between rational subjects, these two principles effectively present to us all of the

[428] Here again, Kant's presentation of the three "maxims of common understanding" mirrors his presentation of the various formulas of the moral law.

implications of both the FUL and the FH, only in a more systematic fashion. In this sense, as Wood has emphasized, it is reasonable to focus on the FA as the most complete presentation of the moral law.[429] But, contrary to Wood, this is precisely because *all three* of Kant's main formulas of the moral law are, in fact, equivalent to one another, *when* they are interpreted as principles for the capacity of practical reason. In other words, the FA is a more complete presentation of the moral law precisely because of how it explicitly relates to the demands of the FUL to the demands of the FH, and not because it is opposed to (or goes fundamentally beyond) these demands in either case.

Indeed, in developing our interpretation of the relationship between the FUL and the FH, we have already noted how close the FA (and FRE) are to those principles, once they are properly understood. But the FA provides us with a systematic understanding of the unity of the FUL and the FH. In doing so, the FA can be seen as referring us back to the connections between autonomy and comprehension we discussed in the last chapter. Here it is interesting to note that Kant phrases the demands of the FA in terms of the command that, "the maxims of one's choice are at the same time comprehended with it in the same volition as universal law". Kant's reference to comprehension in this passage ("*mit begriffen seien*") is often obscured by the manner in which it is translated. For example, in their standard translation of the *Groundwork*, Gregor and Wood render "*mit begriffen seien*" as "be included with". But this translation loses the special connotations of "*begreifen*" for Kant – connotations that are directly relevant to the role that the FA is supposed to play with Kant's moral system. For the FA's demand for the "comprehension" of "one's maxim of choice" and that maxim "as universal law" in a single volition is not simply a demand to "include" both of these in a single volition as (say) a "mere aggregate". Instead, what the FA demands of us is to form a volition that embodies a consciousness of the grounds of one's action, in the context of a system of other ends, in just the manner that practical *comprehension* requires.

In this way, the FA can be seen as expressing the main idea of the last chapter – namely, that autonomy or practical cognition from principles is possible only insofar as one has achieved practical comprehension. Or, in other words, the idea that practical reason's self-consciousness will only be sufficient for its ends if practical reason is conscious of itself as a faculty for practical comprehension. In this, it is the FA which presents the moral law, not just as an abstract formal principle, or as a collection of ends in themselves, but rather as a *system* of both. Thus, for Kant, it is the FA that provides us with our fullest *comprehension* of the moral law – namely, as the principle of practical reason as the capacity for autonomous self-comprehension. In this way, the role of the FA in Kant's account of the moral law is precisely to help us comprehend the

[429] See (Wood 1999; 2007).

moral law in its full systematic glory. As such, it seems to me to be no accident that Kant's statement of the FA itself makes use of precisely this term.[430]

In thinking about the role of the FA (and the moral law more generally) as the principle of the activity of active self-comprehension or autonomous self-realization, it is helpful, I think, to compare Kant's views about these issues with those we find in the early Marx ("Comments on James Mill"):

> 1) In my production I would have objectified my individuality, its specific character, and therefore enjoyed not only an individual manifestation of my life during the activity, but also when looking at the object I would have the individual pleasure of knowing my personality to be objective, visible to the senses and hence a power beyond all doubt. 2) In your enjoyment or use of my product I would have the direct enjoyment both of being conscious of having satisfied a human need by my work, that is, of having objectified man's essential nature [*Wesen*], and of having thus created an object corresponding to the needs of another man's essential nature. 3) I would have been for you the mediator [*der Mittler*] between you and the species, and therefore would become recognized and felt by you yourself as a completion [*Ergänzung*] of your own essential nature and as a necessary part of yourself, and consequently would know myself to be confirmed both in your thought and your love. 4) In the individual expression of my life I would have directly created your expression of your life, and therefore in my individual activity I would have directly *confirmed* and *realized* my true nature, my human nature, my communal nature. (G 1.3)

In his insightful discussion of this passage, (Kandiyali 2020) reads it as presenting the core of an account of unalienated productive activity on which, "fully unalienated production can be defined as work that: 1) involves self-realization, i.e., the exercise, development and manifestation of our individual powers; 2) satisfies another's need; 3) is conducted with the intention of satisfying another's needs; 4) is used and appreciated by that other; and 5) is performed freely." As such, at least on Kandiyali's interpretation of him, Marx presents us with a picture of how unalienated productive activities are possible that shares several important elements with the account of autonomous rationality we have been extracting from Kant. First, Marx shares with Kant the fundamentally Aristotelian assumption that the proper use of an individual's powers must involve the "self-realization" of those powers through their "exercise, development, and manifestation". But, second, much as we have seen above in our discussion of the FH, this activity of self-realization necessarily for Marx takes on an intersubjective dimension – aiming, not merely at self-realization for its own sake, but rather at self-realization insofar as this serves equally to "satisfy the needs" of other rational beings like oneself. Finally,

[430] Say something about "distributed unity" and systematicity here?

the activities in question must be "appreciated by that other" as having this character – that is, it is not merely that each of us wills that he or she act so as to satisfy the needs of others through our own self-realization – rather, each of us wills that everyone act in such a fashion, in accordance with the ideal of universal rational legislation. Only insofar as this is true, for the young Marx, can we speak of a form of genuinely unalienated production. In this sense, for the young Marx – much as for Kant – the full realization of our powers for autonomous agency in the world requires an intersubjectively shared practical comprehension of the goodness of the activities of these capacities as serving the common realization of all of our rational powers in the world. In this sense, as we saw in the last chapter, while the heart of Kant's account of the autonomy of reason lies in his understandings of reason's *internal* self-legislation – the full realization of reason's aims requires that reason goes beyond this to impose its principles and form on the social and physical world. For only insofar as it accomplishes this, will reason have fully realized its aim of autonomous practical comprehension through the reciprocal relations between individual rational agents.

g. The Unity of the (Theoretical and Practical) Principles of Reason

We will return to this connection between Kant's conception of reason and the possibility of critiquing social and political institutions on grounds of irrationality in the conclusion to Part Two below. But for now, I want to complete our discussion of these principles by bringing it together with our previous discussion of the principles of reason in its theoretical use. As should hopefully be clear, we are now well-placed to see how all of these principles follow from the basic idea of reason as an autonomous faculty for theoretical and practical comprehension. In particular, as we discussed above, it follows from Kant's basic conception of a cognitive faculty that any such faculty's aims can be characterized in three interrelated ways – that is, (i) in terms of the *subject* of cognition's representations and activities , (ii) in terms of the *objects* that the faculty ideally cognizes, and (iii) in terms of the *relationship between the subject and these objects* that the faculty aims to achieve. In order to arrive at a systematic presentation of Kant's various formulations of reason's principle, we can pair this tripartite division of perspectives on a faculty's fundamental principle with two further ideas: First, the idea of reason as the faculty for comprehension or cognition from absolute principles, and second, the fundamental distinction between theoretical and practical forms of cognition. For once we do so, we can see each of Kant's formulations of reason's fundamental principle as characterizing what theoretical or practical comprehension requires, from one of these perspectives..

For example, if we adopt the first perspective and focus on the internal relations between representations that comprehension requires, we will quickly be led to the ideas expressed, in turn, by the "logical maxim" in the theoretical case and the Formula of Universal Law in the practical one. For, as we have seen, the "logical maxim" simply characterizes the relationships that must obtain between our various representations

(considered as such) that is required by theoretical reason's aim of theoretical comprehension, insofar as this is possible for us. And similarly, the FUL characterizes the internal structure of the will which is required if practical comprehension of our actions (or the autonomous use of practical reason) is to be possible. Finally, the same basic idea is expressed by the first "maxim of common understanding", only in a less demanding form that makes this idea compatible with the essential limits of both theoretical and practical reason.

In a similar vein, if we approach the nature of theoretical comprehension from the question of what the *objects* of such comprehension must be like, we quickly arrive at the idea that we must assume that these objects comply with what Kant calls the "supreme principle" of theoretical reason – that is, that the objects of ideal comprehension must obey a (highly indeterminate) version of the Principle of Sufficient Reason. In the three ideas of theoretical reason – the ideas of the soul, the world whole, and God – we then see the three forms that this general requirement on the objects of theoretical comprehension takes for creatures who possess our three forms of relational judgment. In the practical domain, we have similarly seen that the Formula of Humanity articulates a basic constraint on the objects of practical comprehension – namely, that any such system of "practical objects" or ends must treat all rational nature as an end in itself. As we discuss in more det, this general idea can be further developed through bringing it together with the three ideas of theoretical reason and the practical significance of happiness so as to form our conception of the highest good, which provides us with our most systematic picture of what the objects of ideal theoretical and practical comprehension must be like.[431] Finally, the same idea is expressed in a more general way in the second of Kant's maxims of common understanding – the "maxim of enlarged thought" – which expresses the manner in which the "objectivity" involved in comprehension is connected with a certain sort of ideal intersubjectivity.

These various principles are then give their most systematic expression when we turn to the third of these perspectives on the principle of reason. Here it is notable that we do not find any principle in the theoretical domain that fits neatly into this category. But this, in some sense, is no surprise – since Kant's view is that it is strictly speaking impossible for us to achieve the sort of relationship to the objects of theoretical cognition which ideal theoretical comprehension requires. As we have seen, things are much more promising when we turn to the practical point of view. For here we do have a grasp of the relationship that ideal practical comprehension requires between practical subject and practical object – and that this relationship obtains precisely insofar as one's choice of action is determined by the autonomous exercise of practical reason. This idea, of course, is what is expressed in the Formula of Autonomy. But we also find this idea in a more explicitly systematic form in the Formula of the Realm of Ends, which makes explicit the systematic

[431] Again, compare Guyer(2000) on this point.

relationships between all rational beings (as both the subjects and the objects of practical cognition) which genuine practical comprehension requires. Finally, we find the same basic idea expressed in a manner which fits both the theoretical and the practical cases in the third maxim of common understanding, whose demand for "consistency" is, as we saw, best understood in terms of the demand for a certain sort of systematic unity in our theoretical and practical cognition.

Taking all of this together, we are now in a position to see why all of these various formulas actually express a single fundamental idea – namely, the idea of reason as an autonomous faculty for comprehension – from a variety of different perspectives and with respect to a variety of different forms that such comprehension can take in us. We can summarize these results in the form of the following "table of principles of reason":[432]

	Principles of Theoretical Reason	**Principles of Practical Reason**	**Maxims of Common Understanding**
Subject-Focused	The "Logical Maxim"	The Formula of Universal Law (The Formula of the Law of Nature)	The Maxim of the Universal Use of Reason
Object-Focused	The "Supreme Principle" (The Ideas of Theoretical Reason)	The Formula of Humanity (The Highest Good)	The Maxim of Enlarged Thought
Complete System of Subjects and Objects		The Formula of Autonomy (The Formula of the Realm of Ends)	The Maxim of Consistency in Thought

g. The Ultimate Object of Human Comprehension: The Highest Good

As this table indicates, Kant's conception of the principles, not just of practical reason, but also of theoretical reason as well, is much more systematic than it might often appear. As we've already alluded to above,

[432] Compare the table in (A. Cohen 2014), but note that this (crucially) neglects the relationship between what I am calling here "principles of theoretical reason" and the other two categories. This neglect, it seems to me, leads Cohen to overstate the degree to which the "maxims of common understanding" are derived from practical *as opposed to* theoretical principles – and so leads Cohen to a conception of the unity of reason as a faculty which over-emphasizes the practical side of this unity over the theoretical.

another way to appreciate the unity of these principles is to see how the idea of the most complete object of practical reason – what Kant calls the "highest good" – provides us with a unified conception of the proper object, not just of practical comprehension or wisdom, but also of theoretical comprehension as well.[433] The highest good, for Kant, is the idea of a good that is complete along all relevant dimensions. Thus, for Kant, the highest good involves – not merely the idea of everyone freely willing the good – but also the idea of a natural world that is systematically related to these wills such that (i) what they will is ultimately made actual and (ii) in particular, happiness is apportioned to agents in accordance with their degree of moral worth.[434] It is this, according to Kant, which forms the proper object of "wisdom" – or, in other words, of the sort of "practical comprehension" which represents the most perfect form of practical cognition we are capable of as human beings. (5:108, 5:130-1)

For Kant the fact that the highest good involves both morality and happiness, in their proper systematic relationship, follows from the fact that we are both rational and sensible beings. Thus, while Kant insists that rationality demands that we treat the value of happiness as conditional on its relationship with morality (and so its relationship with pure practical reason), he does not deny that all of us necessarily have a practical concern for our own happiness. As a result, although the value of happiness is merely conditional for Kant, and in particular is conditional on whether this happiness is deserved, agents like us will necessary experience a lack of deserved happiness as a way in which the world is less good that it ought to be.

Thus, happiness is far from irrelevant to Kant's conception of the good. Rather, as we have already discussed, happiness plays a crucial role in *making determinate* the abstract demands of the moral law. In other words, to understand the implications of the moral law for sensible creatures like ourselves we need to take into account the basic practical significance of happiness for such creatures. Once again, this significance will always be qualified by the relationship between the happiness in question and the moral law. But nonetheless, we cannot make determinate sense of what the moral law demands of us without recognizing the importance that creatures like us necessarily attach to our own happiness. It is this sort of a picture of the good – one which begins with the abstract principles of pure practical reason – but then determines these principles in accordance with the practical significance of happiness for creatures like us – that reaches its most complete form in the idea of the highest good.[435]

This idea is particularly important for our purposes here because Kant claims that the idea of the highest good represents, in some sense, the highest end of reason. This is sometimes taken to indicate that reason is,

[433] See (Gardner 1999; Fugate 2014; Guyer 2005; Wood 1999).
[434] See (Engstrom 1992; Reath 2006).
[435] See also (Guyer 2000).

for Kant, fundamentally concerned with practical matters *as opposed to* (say) cognition from principles or comprehension. But, as we have already seen, this way of describing the relationship between theoretical and practical reason assumes a (merely theoretical) conception of cognition that Kant himself would reject. Indeed, as we have just seen, the "more practical" characterizations that Kant provides of reason's fundamental activity – e.g. in terms of autonomy – are ultimately very closely related to the "more cognitive" characterization he provides of it – e.g. in terms of cognition from principles or comprehension.

Thus, to do justice to the idea of reason as fundamentally autonomous, we do not need to reject the idea of reason as fundamentally a capacity for comprehension or what Kant calls "practical wisdom". Rather, to fully grasp either of these characterizations, we need to understand both *in the context of the other*. And the same is also true with respect to Kant's claim that the highest good represents the ultimate end or object of reason. That is, this claim is best understood – not merely as a claim about reason in its "more practical" guise – but also as a claim about reason under a "more cognitive" guise as well. In this way, this claim is best understood – not as asserting that reason is fundamentally practical *as opposed to* cognitive – but rather as another way of understanding the underlying unity of reason's practical and cognitive dimensions.

To see this, we need to connect Kant's claims about the highest good with our discussion of the ideas of reason in the previous chapter. There, we argued that these ideas are best understood as providing us abstract models for thinking about what a fitting object of ideal or complete theoretical comprehension would have to be like. What I want to argue now is that the idea of the highest good can be seen as a further development of this basic line of thought. In short, I want to argue that the role of the highest good within Kant's system is to provide us with our best grasp of what the *object of complete theoretical and practical comprehension* would have to be like in order for such comprehension to be possible.

To do so, let's begin with Kant's explicit discussion of the highest good in the second *Critique*. There Kant writes that practical reason, "seeks the unconditioned totality of the object of pure practical reason, under the name of the highest good." (5:108) In this way, the idea of the highest good goes beyond the idea of the good will as something that possesses unconditional value. In particular, while the good will is the sole unconditioned good, according to Kant, it does not exhaust the "totality" of the good *qua* the object of practical reason. Instead, in order to arrive at the idea of this "totality" – which Kant calls the highest good – we must place the unconditional value of the good will within a larger system of conditional goods which encompasses all other forms of value and their various relations of dependence. Only by doing this, we will arrive at the idea of the "unconditioned totality of the object of pure practical reason".[436]

[436] Note the two notions of the unconditioned at work. Ascending and descending conditions.

In this way, the idea of the highest good does not just provide us with a picture of the ultimate end that practical reason wills to be actual – it is also provides us with an abstract model for thinking about the object of ideal or complete practical comprehension. Thus, the idea of the highest good – much like the ideas of theoretical reason, on a more limited scale – can be regarded as providing us with characterization of what the object of practical comprehension must be like if complete practical comprehension is to be possible. It is this, again, which makes the highest good the proper object of genuine practical wisdom for Kant. (5:130-1, cf. 5:108)

But, the idea of the highest good does not merely provide us with a grasp of the ideal object of *practical* comprehension – at the same time it provides us with a deeper understanding of the ideal object of *theoretical* comprehension as well. In particular, as we saw above, as long as we limit ourselves to a purely theoretical perspective, our best grasp of what this object is like is provided by the three ideas of theoretical reason – namely, the soul, the world, and God. But, as we discussed there, our purely theoretical grasp of the nature of these things is so indeterminate that we cannot be said to have any sort of theoretical cognition of them.

Fortunately for theoretical reason's aims, in addition to giving us a grasp of what the ideal object of *practical* comprehension must be like, the idea of the highest good also brings these three ideas into a systematic relationship with one another and, by doing so, provides us with a richer and more determinate grasp of their objects. For the essence of the idea of the highest good lies precisely in a system of relationships between these three ideas. As understood by Kant, the idea of highest good consists precisely in the idea of (i) God establishing certain systematic relations between (ii) immortal rational subjects and (iii) the world of appearances. Thus, in thinking of God, the soul, and the world in the terms required by the highest good, we not only more fully determinate our *practical* conception of what ought to be, we also more fully determine our *theoretical* conception of what is – albeit on distinctively practical grounds. Thus, as Kant stresses, the comprehension or "wisdom" that our grasp of the highest good brings with it has both a theoretical and a practical dimension: "*wisdom* theoretically regarded signifies *the knowledge of the highest good* and practically *the uniformity of the will to the highest good* (5:130-1).

In this way, the idea of the highest good provides us with, not just our best grasp of what the object of ideal practical comprehension must be like, but also a richer grasp of what the complete object of ideal theoretical *and* practical comprehension must be like as well. In other words, the highest good provides us with our best grasp of the system of objects (both theoretical and practical) which would satisfy reason's drive for maximally systematic and unified comprehension in both the theoretical and practical domains. As Gardner points out, in virtue of this fact, it is the idea of the highest good that best "satisfies reason's 'architectonic'

interest".[437] That is, it is the idea of the highest good that provides us with our most fundamental model for thinking about the objects of both theoretical and practical comprehension. In this sense, the highest good truly is the highest idea or end of reason in *both* its theoretical and its practical manifestations.[438]

[437] (Gardner 1999)

[438] Compare (Fugate 2014): "Presumably, in view of what we saw with the previous two meanings, this unity of reason would presuppose an idea by means of which the unity of reason in the service of the understanding (theoretical) and the unity of reason commanding in the moral law (practical) are somehow brought into an overarching systematic unity of all possible cognitions of whatever kind (wisdom) ... this idea is precisely that of the highest good as the *bonum consummatum*."