Introduction

Eric Watkins

Our conception of ourselves today as free and equal persons who are capable of acting rationally and autonomously in both practical (moral) and theoretical (scientific) contexts does not reflect an everlasting truth, but rather emerged as a hard-won conquest within a particular historical context. The story of the emergence of this self-conception is long and multifaceted, but one particularly crucial moment occurred in the European Enlightenment when the socio-economic and political structures of the *ancien régime* came under attack by an ambitious and increasingly literate middle class that was pressing its interests against established powers, even against so-called enlightened and benevolent despots, such as Frederick II of Prussia. Underlying the ensuing upheavals in the then dominant social, political, religious, and economic structures were various intellectual developments at the time, which played a central role in the agents' self-professed understanding of what was driving change, the ways in which they formulated and argued their positions, and how they understood themselves.

Many early proponents of the Enlightenment made their case for putting power in the hands of such agents by noting the promise of the technological advances that were possible in the wake of the Scientific Revolution and its implementation in solving local problems. But especially a second generation of advocates advanced a different line of argument, one centered on the authority of human reason as a universally shared capacity, regardless of a person's social rank, position, and religion. This strategy proved in many ways richer and more powerful, since it undergirded and applied to a broader range of activities and contexts and had nearly universal appeal as a result. Arguably the most influential and most profound thinker to articulate this line of argument was Immanuel Kant. While his positions on the nature of reality (Transcendental Idealism), the limits of our knowledge of it (Epistemic Humility), and the fundamental principle of morality (the Categorical Imperative) have – with good reason – received

the lion's share of scholars' attention, it is his view of the nature of reason itself that is even more fundamental. Particularly important is Kant's thesis of the unity of reason, the idea that both theoretical and practical reason function according to the very same principles within a free and autonomous agent. For this faculty has implications for what a person's most valuable capacities are, and, consequently, for the various ways in which we can understand our place and life-projects in the world.

That a people could rationally will laws that would be both scrutinized by all and binding on all is a powerful idea that Rousseau popularized, one that Kant then articulated more fully in his metaphysics and in his moral and political philosophy (with its emphasis on the public use of reason and its authority). In religion, Kant advocated the use of reason both to criticize superstition (so as to avoid "enthusiasm") and to lay bare the fundamental rationality of the "invisible church" that would unite all people in a "realm [*Reich*] of ends." In such a realm all persons, in virtue of their rationality, are treated not as things with a price, but as ends in themselves worthy of unconditional respect, precisely because of their capacity for autonomous agency. In these and numerous other ways, Kant articulated a powerful and enduring conception of what it means to be a person acting in a complex and ever-changing world, a conception that was, to be sure, criticized and modified in various ways by his successors, but one that was accepted in many respects. For example, Kant's emphasis on the fundamental dignity of persons is reflected in modern constitutions (e.g., in Germany's *Grundgesetz*), which reveals that his conception continues to be an attractive view of how we understand ourselves most fundamentally today.

The present volume investigates three different aspects of Kant's conception of agency: autonomy, freedom, and personality. The first part is devoted to autonomy and how agency relates to it. The second considers freedom and its role in Kant's account of agency. The third focuses on Kant's conception of persons and how persons are agents. The volume concludes with a synoptic vision of Kant's conception of "the end of all things."

In the first chapter, "The Unconditioned Goodness of the Good Will," Eric Watkins considers what it means to assert, as Kant does in the first section of the *Groundwork for the Metaphysics of Morals*, that the good will alone is an *unconditioned* good and that all other goods are *conditioned* by its goodness. He begins by distinguishing different kinds of goodness (e.g., intrinsic vs. extrinsic, conditioned vs. unconditioned, good with vs. without limitation), before clarifying what Kant means by a condition in his theoretical philosophy (metaphysical dependence that is also explanatory), and using that conception to account for what unconditioned goodness is.

He then explains how the goodness of the good will is related to the kind of universal legislation that is fundamental to Kant's conception of autonomy, and to Kant's claim at the very end of the *Groundwork* that there are limits to what we can comprehend about the unconditioned necessity of the laws that we autonomously legislate to ourselves.

In the second chapter, "Universal Law," Allen Wood considers several of Kant's formulas of the Categorical Imperative, which express in different ways how we should behave if we are to act morally and autonomously. He argues that the so-called Universal Law and Law of Nature formulas of the Categorical Imperative neither can be, nor are intended by Kant to serve as, universal criteria for distinguishing right from wrong on any given occasion. Were they intended as such, they would be subject to both false negatives and false positives, as critics have charged. Instead, Wood claims, Kant uses these formulas, especially that of the Law of Nature, to expose the illegitimacy of the justifications that one might offer in defense of maxims that would (improperly) exempt one from the moral law. Wood supports his claim by discussing several of the examples that Kant employs, showing how this more modest intent determines what maxims are selected for discussion, what questions are asked about these maxims, the purpose for which the agent asks these questions, and even the specific moral defects and virtues that the agent displays in using the formulas.

Stephen Engstrom's contribution in the third chapter of this volume, "Understanding Autonomy: Form and Content of Practical Knowledge," explains crucial features of Kant's conception of autonomy in terms of his conception of practical knowledge. In particular, Engstrom is concerned to show how a proper understanding of practical knowledge can allow one to address two objections commonly raised against Kant's doctrine of autonomy, namely that if we (as rational beings endowed with a will) are the source of our moral obligations, then they cannot be at once necessary and contentful. They cannot be necessary, since we could, it seems, rescind them at will, and even if a purely formal law, like the law of contradiction, might be necessary, it cannot have any content, since it is purely formal. Given that Kant conceives of the will that legislates for itself as practical reason and of practical reason as a capacity for practical knowledge, Engstrom then provides a detailed description of both the form and the content of practical knowledge as involving acts of self-legislation. In light of Engstrom's analysis, our moral obligations have a necessary content because the form of practical rational knowledge is not only *legislative* (and thus necessary), but also *self*-legislative (and thus contentful insofar as the self necessarily brings a content into its own knowledge).

In Chapter 4, "The Principle of Autonomy in Kant's Moral Theory: Its Rise and Fall," Pauline Kleingeld notes that Kant's Principle of Autonomy, which played a central role both in the *Groundwork for the Metaphysics of Morals* and in the *Critique of Practical Reason*, had all but disappeared by the time of the *Metaphysics of Morals*. She argues that its disappearance is due to significant changes in Kant's political philosophy. That is, whereas the notion of legislation, or lawgiving, that Kant accepted in the mid 1780s does not require any actual consent – genuine universality is sufficient for a law to be just – in the *Metaphysics of Morals* and in other works in the 1790s he added the further condition that laws must be given by the citizens themselves, through their representatives in parliament. With this further condition, the analogy that Kant saw between his political and moral philosophy in the mid 1780s no longer obtained, and the Principle of Autonomy, which is firmly based on that analogy, is no longer suitable for its original purpose.

The second part of the volume focuses on freedom and its role in agency. Chapter 5, "Evil and Practical Reason," by Lucy Allais, explores the relation between Kant's account of practical reason as autonomy, the idea of freedom in his political philosophy, and his account of the innate evil in human nature. It offers a secular account of Kant's thesis of innate evil, understood in terms of our being imperfect creatures who come into a world in which we are unavoidably situated in relations of current and historical systematic injustice that taint our moral options. In particular, seeing yourself as an agent (someone who acts for reasons) involves seeing your actions as governed by the constraint of respecting the humanity of others, which suggests that there is internal pressure to see yourself as having an ordered will of a certain sort (to interpret yourself as basically good), since this is part of what it is to see yourself as a rational agent who acts for reasons. This suggests further a picture of rational agency which contains a mechanism by which self-deception is likely to arise in circumstances of systematic injustice and to take the form of dehumanizing others. If we need to see ourselves as good to some degree in order to see ourselves as agents, but we find ourselves in circumstances in which we know we are going to fail to be good, we may be liable to despair, and thus be under internal psychological pressure to dehumanize others so as to avoid confronting the ways in which we are implicated in injustice and domination. The role of affective attitudes such as forgiveness and trust is that they enable us to avoid despair by providing an optimistic perspective on our future willing that may be necessary for our properly seeing ourselves as agents.

Chapter 6, "Freedom as a Postulate," by Marcus Willaschek, focuses on solving two puzzles that arise concerning Kant's views on how freedom

could be a postulate. First, why does Kant not provide an argument for the postulate of freedom in the *Critique of Practical Reason*'s Dialectic, just as he does for those of God and the immortality of the soul? Second, how can freedom be a *postulate* if it is *proved* on the basis of Kant's famous "fact of reason"? Willaschek provides a detailed reconstruction of Kant's "fact of reason" argument, which shows that his missing argument for freedom *as a postulate* can be found in the Analytic and that this does not undermine his claim that the appropriate doxastic attitude toward freedom is belief (as opposed to knowledge). But Willaschek also draws a more general lesson from Kant's position, namely that Kant's conception of a postulate of practical reason is both broader than has been thought and also more attractive as a contemporary position than commentators have acknowledged.

In the seventh chapter, "The Struggle for Freedom: Freedom of Will in Kant and Reinhold," Paul Guyer argues that throughout his career Kant was committed both to the distinction between *Wille* and *Willkür* (even if not in those exact words) and to the thesis of the freedom of *Willkür* to choose between good and evil (again, even if not always in those very words). He then shows why Kant sometimes suggested otherwise, but argues that his fundamental reason for insisting on the freedom of *Willkür* is compelling. What's more, Guyer suggests that we should not take Kant's repudiation of the *definition* of freedom of *Willkür* as the ability to choose either the moral law or its subordination to self-love, to repudiate either the difference between *Wille* and *Willkür* or Kant's commitment to the freedom of *Willkür*. In this way Guyer is able to diagnose and rectify prominent misconstruals of Kant's position by Reinhold and others.

In Chapter 8, "The Practice of Self-Consciousness: Kant on Nature, Freedom, and Morality," Dieter Sturma argues that Kant's solution to the problem of freedom and natural determinism in the third antinomy, which is based on his conception of a causality through freedom, is not acceptable under the terms of contemporary systematic philosophy. The primary object of criticism is Kant's presupposition of a dualistic theoretical approach and the associated two-worlds view of the empirical and the noumenal. However, in his conception of freedom and agency, Kant is not necessarily obliged, Sturma argues, to accept a strong interpretation of the two-worlds view. Instead, his critical philosophy is systematically determined by two orders: the *realm of causes* and the *space of moral reasons*, which Kant has in mind when he invokes the image of the "starry heavens above and the moral law within." Sturma argues further that, at the end of the *Critique of Practical Reason*, Kant emphasizes that these two orders are closely interwoven with the self-consciousness of the person, which excludes an ontological dualism. He then points to a number of Kant's reflections and hints

about the practice of self-consciousness from which – as an unofficial doctrine – a conception of freedom and agency can be derived that is not committed to a strong version of the two-worlds view and that is compatible with his concept of autonomy. In short, according to Kant, the life of a person is characterized by the ability to set empirical conditions against rational constraints, since persons are beings who can respond to reasons as well as generalize, differentiate, and act due to reasons. The laws of nature and the moral laws thus have to satisfy different standards of validity and, accordingly, express themselves in different kinds of objectivity.

The third part of the volume considers Kant's conception of a person and the central role that agency plays in it. In Chapter 9, "Kant's Multiple Concepts of Person," Béatrice Longuenesse argues that in the course of his criticism, in the Third Paralogism, of the rationalist derivation of the concept of a person from the mere use of 'I' in 'I think,' Kant offers resources for developing an alternative notion of person. This is the notion of a person as an embodied entity endowed with unity of apperception and with the capacity for moral accountability. This is not, however, the notion of person Kant himself endorses at the end of his criticism of the paralogism of personhood. Rather, there Kant claims that the rationalist notion of person that was the target of his criticism can remain, albeit on behalf of the practical rather than the theoretical use of reason. Longuenesse offers an analysis of this surprising about-face on Kant's part, comparing it to his own pre-critical attempt to derive a notion of person from the mere analysis of our use of 'I' in 'I think' and 'I do.' She then argues that in preserving a rationalist notion of person for practical use, Kant is prey to his own paralogism, which she calls a paralogism of pure practical reason. Finally, she suggests that the empirical notion of person one might have expected to emerge from Kant's criticism of the third Paralogism can be seen as an ancestor of the notion of person Harry Frankfurt offers in "Freedom of the Will and the Concept of a Person." Kant's concept differs from Frankfurt's, however, in offering a criterion for the second-order assessment of one's first-order volitions: the categorical imperative of morality.

In Chapter 10, "We Are Not Alone: A Place for Animals in Kant's Ethics," Barbara Herman presents a novel account of Kant's position on our moral obligations with respect to animals. Specifically, Herman argues that one should not read the famous "Amphiboly" passage from the *Metaphysics of Morals* – where Kant speaks directly to our duties to self with respect to animals – as arguing that we should avoid violent and cruel treatment of animals only out of moral concern for maintaining the affective system that supports us in our duties to human beings. Instead, she

draws on the way in which we come to know ourselves as embodied beings in and through the life activity of animals, through our experience of their movements and activities as well as of their pain and suffering, and that this experience helps us to understand more fully what is at stake in paying attention to one's ends. Herman takes this connection to animal life and bodies to suggest reading the Amphiboly argument as follows: if there were no duty with respect to animals, if we were allowed to ignore and override their pain, in permitting cruelty to animals, we would thereby treat *our* natural unconditional response to suffering as, morally speaking, conditional, which runs counter to basic tenets of Kant's moral theory.

In "The Dynamism of Reason in Kant and Hegel," the eleventh chapter in the volume, Robert Pippin focuses on the sense in which Kant, especially and increasingly in (and around) the third *Critique*, came to understand that reason not only is responsible for a spontaneous activity, but also is a purposive, self-actualizing, self-determining, and teleologically structured faculty of agents. It thus has an irreducible practical and productive character. This fits with Kant's characterization of reason as having needs and desires, that is, a fundamentally conative character involving a practical necessity (even if its desires can never be fully satisfied theoretically). Pippin then shows that Hegel picks up on and develops these features of reason further throughout his own philosophical system (including the *Science of Logic*) by understanding reason's understanding of itself and its own activity as fundamentally developmental even as he introduces non-Kantian elements, such as a dynamic notion of determinate negation.

The volume concludes with "Once Again: The End of All Things," by Karl Ameriks. "The End of All Things," is, appropriately enough, one of the last things that Kant published, and it is still generally regarded as one of his most mysterious works. Ameriks argues that, like many of the late essays, it has a complex political–theological subtext, while also being one of the few pieces in which Kant tries to connect, all at once, the implications of his notions of the person and agency with his complex metaphysical doctrines of the transcendental ideality of time and the idea of the highest good. Ameriks focuses especially on the importance of moral considerations to determine the most appropriate attitude toward our ultimate fate and on the sense in which we ought to take our immortality. In raising these issues, Ameriks provides a fitting concluding perspective on what is still alive in the final phase of Kant's publications.

PART I

Autonomy

CHAPTER 1

The Unconditioned Goodness of the Good Will

Eric Watkins

1 INTRODUCTION

In the first sentence of the first section of the *Groundwork of the Metaphysics of Morals*, Kant famously claims that the good will alone is good without limitation (*ohne Einschränkung*). In the pages that follow, it then becomes clear that Kant also maintains that the good will alone is an unconditioned good.[1] Unfortunately, it is not immediately obvious exactly how these two claims are related.[2] But recently much attention has been devoted especially to understanding the second claim and to reconstructing Kant's arguments for the claim offered in support of it, namely that all other things we might naturally consider good, such as power, money, health, and happiness, are in fact only conditioned goods, since a good will is a condition of their goodness. Now Kant's characterization of the goodness of the good will as unconditioned is supposed to be expressive of the standpoint of "common rational moral cognition" (G 4:393), but it also displays striking parallels with claims he makes in his theoretical philosophy. For example, Kant characterizes the objects of traditional metaphysics – God, freedom, and the soul – as unconditioned and he places them at the center of the drama that then unfolds throughout the *Critique of Pure Reason*.

[1] Note that in the following, I speak of something being an unconditio*ned* good, as opposed to it being unconditio*nally* good for two reasons. First, Kant typically does not use "*unbedingt*" as an adverb, so I am simply retaining his own use of the term. Second, there is a philosophical difference between the property of being an unconditioned good and having the property of being good in an unconditional way, since being an unconditioned good entails having a specific kind of goodness, whereas being good in an unconditional way is a specific way in which one has a potentially ordinary kind of goodness. A similar distinction can be made with respect to intrinsicality. The property of being in the room with President Obama is a property that things can have either extrinsically, as I do when I visit the White House in his presence, or intrinsically, as President Obama has whenever he is in a room.

[2] Given what Kant explicitly says in the first few sentences of the *Groundwork*, one might think that they are equivalent. As will become clear below, I agree with Karl Ameriks (*Interpreting Kant's Critiques*, Oxford: Oxford University Press, 2003, p. 198) that being good without limitation follows from being an unconditioned good (and not vice versa).

And these parallels are no accident. Instead, they are to be explained by the fact that Kant conceives of reason in general as a spontaneous faculty that seeks not only the conditions of whatever is conditioned but also the totality of such conditions and thus the unconditioned. It is this conception of reason that results in and provides unity to the uses that Kant makes of the notion of the unconditioned in both his theoretical and his practical philosophy.

Once one notices these parallels and the generic concept of reason that underlies them, one can see more clearly that in the context of his practical philosophy, Kant's interest in the notions of a condition and of the unconditioned extends beyond simply noting the unconditioned goodness of the good will. For example, in the *Critique of Practical Reason*, Kant argues that, in a manner analogous to theoretical reason, practical reason "seeks the unconditioned for the practically conditioned [...] it seeks the unconditioned totality of the object of pure practical reason, under the name of the *highest* good" (KpV 5:108). In determining the concept of the highest good as the proper object of practical reason, Kant identifies virtue as the "supreme condition of whatever can even seem to us desirable," that is, as "the condition which is itself unconditioned" (KpV 5:110). These passages from the second *Critique* help us to see that Kant's claim in the *Groundwork* about the unconditioned goodness of the good will is but one of many that involve the concepts of the condition and the unconditioned in his practical philosophy.

However, we find other uses of these notions in Kant's practical philosophy that do not fit into this larger picture quite as neatly, at least at first glance. For example, Kant refers to the Categorical Imperative as an "unconditioned command" (G 4:420), as being "limited by no condition" (G 4:416), and as containing "no condition to which it would be limited" (G 4:421).[3] Kant is making a related point when he says that duty is the "unconditioned necessity of action; it must therefore hold for all rational beings" (G 4:425). At least part of what Kant means to be asserting by claiming that the Categorical Imperative is an unconditioned command is that it has normative force regardless of one's desires, independent of any contingent features of human nature, and in virtue of one's rationality, but these points are clearly distinct from his claims about the unconditioned goodness of the good will and the highest good.

[3] In a similar vein, Kant claims that "only law brings with it the concept of an *unconditioned* and objective and hence universally valid *necessity*, and commands are laws that must be obeyed, that is, must be followed even against inclination" (G 4:416).