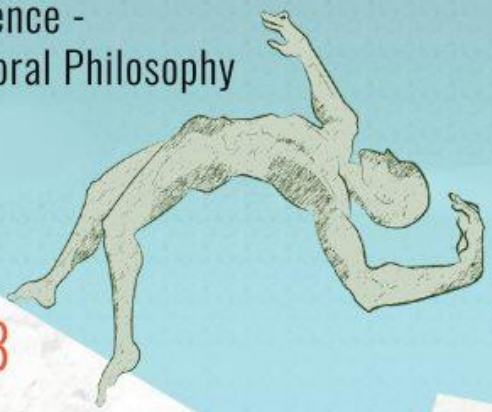
The Rational Human Condition

Kantian Ethics and
Human Existence -
A Study in Moral Philosophy



VOLUME 3

World Philosophy

ROBERT HANNA

NOVA

WORLD PHILOSOPHY

THE RATIONAL HUMAN CONDITION

VOLUME 3

KANTIAN ETHICS AND HUMAN EXISTENCE: A STUDY IN MORAL PHILOSOPHY

WORLD PHILOSOPHY

Cover Art: “The Human Condition Rationalized,” by Otto Paans

The first four hard-copy volumes in THE RATIONAL HUMAN CONDITION series can be found on Nova’s website, **HERE**.

The first four e-books in THE RATIONAL HUMAN CONDITION series can be found on Nova’s website, **HERE**.

WORLD PHILOSOPHY

THE RATIONAL HUMAN CONDITION

VOLUME 3

KANTIAN ETHICS AND HUMAN EXISTENCE: A STUDY IN MORAL PHILOSOPHY

ROBERT HANNA

nova
science publishers
New York

Library of Congress Cataloging-in-Publication Data

ISBN: 978-1-53614-521-2

Published by Nova Science Publishers, Inc. † New York

All five volumes of THE RATIONAL HUMAN CONDITION
are dedicated to the people I love—you know who you are.
But especially Martha and Beth.
And also to all those who helped me with its ideas and arguments.

CONTENTS

"The Human Condition," by Thomas Whitaker/Prison Arts Coalition.

PREFACE

Robert Hanna's THE RATIONAL HUMAN CONDITION is a five-volume book series, including:

- Volume 1. *Preface and General Introduction, Supplementary Essays, and General Bibliography*
- Volume 2. *Deep Freedom and Real Persons: A Study in Metaphysics*
- Volume 3. *Kantian Ethics and Human Existence: A Study in Moral Philosophy*
- Volume 4. *Kant, Agnosticism, and Anarchism: A Theological-Political Treatise*
- Volume 5. *Cognition, Content, and the A Priori: A Study in the Philosophy of Mind and Knowledge*

The fifth volume in the series, *Cognition, Content, and the A Priori*, was published by Oxford University Press in 2015. So, with the present publication of the first four volumes in the series by Nova Science in 2019, all five volumes of THE RATIONAL HUMAN CONDITION are now available in hard-copy and as e-books. All five books share a common aim, which is to work out a true general theory of human rationality in a thoroughly nonideal natural and social world. This philosophical enterprise is what Hanna calls rational anthropology. In the eleventh and most famous of his Theses on Feuerbach, Karl Marx wrote that "philosophers have only interpreted the world in different ways; the point is to change it." Hanna completely agrees with Marx that the ultimate aim of philosophy is to change the world, not merely interpret it. So, Marx and Hanna are both philosophical liberationists: that is, they both believe that philosophy should have radical political implications. But, beyond Marx, Hanna also thinks that the primary aim of philosophy (understood as rational anthropology) and its practices of synoptic reflection,

writing, teaching, and public conversation is to change lives for the better—and ultimately, for the sake of the highest good. Then, and only then, can the human race act upon the world in the right way. The first four volumes of THE RATIONAL HUMAN CONDITION will therefore appeal not only to philosophers, but also to any other philosophically-minded person interested in the intellectual and practical adventure of synoptic, reflective thinking about the nature of our rational, but still ineluctably "human, all-too-human" lives.

A NOTE ON REFERENCES

Throughout the four-volume series THE RATIONAL HUMAN CONDITION, for convenience, I refer to Kant's works infratextually in parentheses. The references include both an abbreviation of the English title and the corresponding volume and page numbers in the standard "Akademie" edition of Kant's works: *Kants gesammelte Schriften*, edited by the Königlich Preussischen (now Deutschen) Akademie der Wissenschaften (Berlin: G. Reimer [now de Gruyter], 1902-). I generally follow the standard English translations, but have occasionally modified them where appropriate. For references to the first *Critique*, I follow the common practice of giving page numbers from the A (1781) and B (1787) German editions only. Here is a list of the relevant abbreviations and English translations:

BL "The Blomberg Logic." In *Immanuel Kant: Lectures on Logic*. Trans. J. M. Young. Cambridge: Cambridge Univ. Press, 1992. Pp. 5-246.

C *Immanuel Kant: Correspondence, 1759-99*. Trans. A. Zweig. Cambridge: Cambridge Univ. Press, 1999.

CF *Conflict of the Faculties*. Trans. M. Gregor. Lincoln, NE: Univ. of Nebraska Press, 1979.

CPJ *Critique of the Power of Judgment*. Trans. P. Guyer and E. Matthews. Cambridge: Cambridge Univ. Press, 2000.

CPR *Critique of Pure Reason*. Trans. P. Guyer and A. Wood. Cambridge: Cambridge Univ. Press, 1997.

CPrR *Critique of Practical Reason*. Trans. M. Gregor. In *Immanuel Kant: Practical Philosophy*. Cambridge: Cambridge Univ. Press, 1996. Pp. 139-271.

DiS "Concerning the Ultimate Ground of the Differentiation of Directions in Space." Trans. D. Walford and R. Meerbote. In *Immanuel Kant: Theoretical Philosophy: 1755-1770*. Cambridge: Cambridge Univ. Press, 1992. Pp. 365-372.

DSS "Dreams of a Spirit-Seer Elucidated by Dreams of Metaphysics." Trans. D. Walford and R. Meerbote. In *Immanuel Kant: Theoretical Philosophy: 1755-1770*. Pp. 301-359.

EAT "The End of All Things." Trans. A. Wood and G. Di Giovanni. In *Immanuel Kant: Religion and Rational Theology*. Cambridge: Cambridge Univ. Press, 1996. Pp. 221-231.

GMM *Groundwork of the Metaphysics of Morals*. Trans. M. Gregor. In *Immanuel Kant: Practical Philosophy*. Pp. 43-108.

ID "On the Form and Principles of the Sensible and Intelligible World (Inaugural Dissertation)." Trans. D. Walford and R. Meerbote. In *Immanuel Kant: Theoretical Philosophy: 1755-1770*. Pp. 373-416.

IUH "Idea for a Universal History with a Cosmopolitan Aim." Trans. A. Wood. In *Immanuel Kant: Anthropology, History, and Eduction*. Cambridge: Cambridge Univ. Press, 2007. Pp. 107-120.

JL "The Jäsche Logic." Trans. J. M. Young. In *Immanuel Kant: Lectures on Logic*. Pp. 519-640.

LE *Immanuel Kant: Lectures on Ethics*. Trans. P. Heath. Cambridge: Cambridge Univ. Press, 1997.

MFNS *Metaphysical Foundations of Natural Science*. Trans. M. Friedman. Cambridge: Cambridge Univ. Press, 2004.

MM *Metaphysics of Morals*. Trans. M. Gregor. In *Immanuel Kant: Practical Philosophy*. Pp. 365-603.

OP *Immanuel Kant: Opus postumum*. Trans. E. Förster and M. Rosen. Cambridge: Cambridge Univ. Press, 1993.

OPA "The Only Possible Argument in Support of a Demonstration of the Existence of God." Trans. D. Walford and R. Meerbote. In *Immanuel Kant: Theoretical Philosophy: 1755-1770*. Pp. 107-201.

OT "What Does It Mean to Orient Oneself in Thinking?" Trans. A. Wood. In *Immanuel Kant: Religion and Rational Theology*. Pp. 7-18.

Prol *Prolegomena to Any Future Metaphysics*. Trans. G. Hatfield. Cambridge: Cambridge Univ. Press, 2004.

PP "Toward Perpetual Peace." Trans. M. Gregor. In *Immanuel Kant: Practical Philosophy*. Pp. 317-351.

Rel *Religion within the Boundaries of Mere Reason.* Trans. A. Wood and G. Di Giovanni. In *Immanuel Kant: Religion and Rational Theology*. Pp. 57-215.

RTL "On a Supposed Right to Lie from Philanthropy." Trans. M. Gregor. In *Immanuel Kant: Practical Philosophy*. Pp. 611-615.

VL "The Vienna Logic," Trans. J. M. Young. In *Immanuel Kant: Lectures on Logic*. Pp. 251-377.

WiE "An Answer to the Question: 'What is Enlightenment?'" Trans. M. Gregor. In *Immanuel Kant: Practical Philosophy*. Pp. 17-22.

Chapter 1

INTRODUCTION: EXISTENTIAL KANTIAN ETHICS AND THE NATURE OF MORALITY

It is impossible to think of anything at all in the world, or even beyond it, that could be considered good without limitation except a **good will**…. A good will is good not because of what it effects or accomplishes, because of its fitness to attain some proposed end, but only because of its volition, that is, it is good in itself and, regarded for itself, is to be valued incomparably higher than all that could be brought about by it in favor of some inclination and indeed, if you will, of the sum of all inclinations…. In the natural constitution of an organized being, that is, one constituted purposively for life, we assume as a principle that there will be found in it no instrument for some end other than what is also most appropriate to that end and best adapted to it. Now in a being that has reason and a will, if the proper end of nature were its *preservation*, its *welfare*, in a word its *happiness*, then nature would have hit upon a very bad arrangement in selecting the reason of the creature to carry out this purpose…. Since reason is not sufficiently competent to guide the will surely with regard to its objects and the satisfaction of all our needs (which it to some extent even multiplies)—an end to which an implanted natural instinct would have led much more certainly; and since reason is nevertheless given to us as a practical faculty, that is, as one that is to influence the *will*; then, where nature has everywhere else gone to work purposively in distributing its capacities, the true vocation of reason must be to produce a will that is good, not perhaps *as a means* to other purposes, but *good in itself*, for which reason was absolutely necessary. This will need not, because of this, be the sole and complete good, but it still must be the highest good and the condition of every other [good], even of all demands for happiness. (*GMM* 4: 393-396, boldfacing in the original)

Purity of Heart Is to Will One Thing …. [T]he person who in truth wills only one thing *can will only the good*, and the person who wills only one thing when he wills the good *can will only the good in truth*.[1]

1.1 *EXISTENTIAL* KANTIAN ETHICS

The version of Kantian ethics that I develop in this book is "existential" in four senses of that much-used (and much-abused) term. First, it is a specifically *anthropocentric*, humane version of Kantian ethics, that takes philosophical anthropology fully seriously for the purposes of ethical theory, and not as an inessential add-on or mere elaboration.[2] Second, it is a *non-reductively naturalistic* and *organicist*[3] Kantian ethics, in that it is fully embedded in the complex dynamic, purposively biological, neurobiological *lives* of rational human animals or real human persons, even though the guiding principles of choice and action in those lives are categorical, non-instrumental imperatives. Third, it is an *applied* and *situated* Kantian ethics, in that it is specifically intended to apply to real-life, real-world, "human, all too human" moral issues under thoroughly nonideal natural and social conditions. And fourth, it is a Kantian ethics that is significantly informed by writings in the post-Kantian tradition of philosophical and literary *Existentialism*—as well, of course, as being significantly informed by recent and contemporary ethical theory.

It might already surprise you that a Kantian ethics could be "existential" in any of those senses, much less in all four of them. Indeed, the fundamental problems with classical Kantian ethics are generally supposed *also* to be fourfold, as follows—

> (i) its excessive formalism,
> (ii) its rigorism (that is, the overstrictness, overgeneralization, and overextension of its moral rules),
> (iii) its lack of direct engagement with actual human beings, their actual psychological motivations, and their actual historical situations, and
> (iv) its extreme moral rationalism.

And looking at it from the other, classical Kantian side, one might also wonder how an existential version of Kantian ethics could ever manage to avoid committing *the naturalistic fallacy*, according to which facts about moral obligation are strictly determined by sense experiential and/or contingent natural (including physical, chemical, biological/evolutionary, neurobiological, or sociobiological) facts, thereby fallaciously reducing the *ought* to the *is*?[4]

So what precisely is it that sets *Existential Kantian Ethics* apart from other ethical theories, including other recent and contemporary versions of Kantian ethics,[5] yet still manages to preserve the unshakeable integrity of classical Kantian absolute, nondenumerably infinite, intrinsic, objective values, as well as the super-strong normative force of the Categorical Imperative and the other classical Kantian a priori moral principles? In order to answer this question, I will need to say something in an introductory way about the distinction between *ethics* and *morality*, and also about the nature of a specifically existential Kantian approach to them both.

1.2 ETHICS, MORALITY, AND EXISTENTIAL KANTIAN ETHICS

As Hegel in the 19th century and also many more recent or contemporary philosophers—perhaps most notably, in the 1970s, Bernard Williams—have correctly noted, it is illuminating to distinguish between "ethics" (aka *Sittlichkeit*) and "morality" (aka *Moralität*).[6] *Ethics* is the larger, more encompassing domain of *values, especially including the highest good(s)*, and *morality*, the domain of *rules, principles, strict normative laws, permissions, and obligations*, is only a proper part of it. On Williams's account, strikingly, morality is "the peculiar institution," alluding of course to John C. Calhoun's notorious description of the American system of slavery prior to the Civil War.[7] By ironically applying this morally uncomplimentary label to *morality itelf*, Williams means that it is nothing but a socially constructed, life-denying, normatively shallow, inherently oppressive, inhumane, and self-perpetuating formal sub-system of *rule-mongering* within our real, fully meaningful, "thick," multi-textured, and all-encompassing "human, all too human" ethical life.[8] Similar critical, skeptical thoughts about morality have been developed by Nietzsche, Michel Foucault, and John Mackie.[9]

But on my sharply different understanding of the ethics vs. morality distinction, morality is the *essence* of ethics. Our ethical life is indeed real, fully meaningful, "thick," multi-textured, "human, all too human," and all-encompassing ethical life. But morality is its all-enabling core. So according to Existential Kantian Ethics, morality is a proper part of ethics *only* in the very special sense in which, for an Aristotelian essentialist theory of the whole-part relation, *the essential structures of wholes are proper parts of them*.[10] In this sense, the proper part structurally guides and pervades the whole.

On my understanding, ethics, with morality as its guiding and pervasive structural essence, is all about what I call *rational minded animals* and *rational normativity*. But I also need to say what I mean by these terms.

By a *minded animal*, as I have noted in the other books in THE RATIONAL HUMAN CONDITION, I mean any living organism with inherent capacities for:

> (i) *consciousness*, that is, a capacity for embodied subjective experience,
> (ii) *intentionality*, that is, a capacity for conscious mental representation and mental directedness to objects, events, processes, facts, acts, other animals, or the subject herself (so in general, a capacity for mental directedness to *intentional targets*), and also for
> (iii) *caring*, that is, a capacity for conscious affect, desiring, and emotion, whether directed to objects, events, processes, facts, acts, other animals, or the subject herself.

And as I have also noted in those other books, over and above consciousness, intentionality, and caring, in some minded animals, there is also a further inherent capacity for

> (iv) *rationality*, that is., a capacity for self-conscious thinking according to principles and with responsiveness to reasons, hence poised for justification, whether logical thinking (including inference and theory-construction) or practical thinking (including deliberation and decision-making).

Rational minded animals are also the same as what, in those other books, I call *real persons*.[11]

By *rational normativity*, in turn, I mean this irreducible two-part fact:

> (i) that all rational minded animals or real persons have aims, commitments, ends, goals, ideals, and values—hence, as rational animals, they are also teleological animals, and
> (ii) that these rational minded animals or real persons naturally treat their aims, commitments, ends, goals, ideals, and values—hence, as rational and teleological animals, they naturally treat these telic targets
> (iia) as rules, principles, or laws for guiding theoretical inquiry and practical enterprises,
> (iib) as reasons for justifying beliefs and intentional actions, and also
> (iic) as standards for critical evaluation and judgment.

Furthermore, rational normativity in this sense can be

> either (i) *instrumental*, that is, conditional, hypothetical, desired for the sake of some further desired end, pragmatic, prudential, or consequence-based,
> or (ii) *non-instrumental*, that is, unconditional, categorical, desired for its own sake as an end-in-itself, non-pragmatic, non-prudential, and obtaining no-matter-what-the-consequences.

As such, norms provide *reasons* for belief, cognition, knowledge, and intentional action, and categorical norms provide *categorical or overriding reasons* for belief and intentional action. Moreover, categorical norms are *fully consistent with* norms that are instrumental, conditional, desired for the sake of other ends, pragmatic, prudential, and obtaining only in virtue of good consequences. Nevertheless, categorical norms are also *strictly underdetermined by* all other sorts of norms, and therefore cannot be assimilated to or replaced by those other sorts of norms. Finally, cutting across all these notions, there are also two importantly distinct kinds of rational normative standards:

> (i) *minimal or nonideal standards*, which specify a "low-bar" set of goals, targets, rules, principles, or laws, below which normatively evaluable activity cannot and does not occur at all, and which therefore jointly constitute a qualifying level of normativity, and
> (ii) *maximal or ideal standards*, which necessarily include and presuppose the (satisfaction of the) minimal, non-ideal, or low-bar standards, but also specify a further "high-bar" set of goals, targets, rules, principles, or laws, below which normatively evaluable activity indeed occurs, but is always more or less imperfect, and in certain relevant respects, bad

> activity, and above which more or less perfected, and in the relevant respects, fully good activity occurs, and which therefore jointly constitute a perfectionist level of normativity.

From all this, I infer four things. First, all rational normativity includes both low-bar or qualifying standards and also high-bar or perfectionist standards. Second, the satisfaction of the high-bar standards necessarily requires the satisfaction of the low-bar standards. Third, the satisfaction of the low-bar standards is not in itself sufficient for the satisfaction of the high-bar standards. And finally, fourth, failing to satisfy the high-bar standards is not in itself sufficient for failing to satisfy the low-bar standards. Collectively, this is what I call *The Two-Dimensional Conception of Rational Normativity*.

Against the backdrop of those conceptions of rational minded animals, or real persons, and rational normativity, I want to say that a specifically *existential* ethics is all about the aims, commitments, goals, ideals, values or ends of the lives of rational and also specifically *human* minded animals, or real and also specifically *human* persons, in a thoroughly nonideal natural and social world. In such a world, as essentially finite, "human, all too human" animals, living lives that include our own inevitable deaths as a necessary limit, each one of our lives is a *fundamental project* that consists in *a search for individual and collective meaning and purpose*. Finding ourselves in such a world, already and always embarked on such a fundamental project, we are naturally presented with values or ends. We naturally desire those ends. Then we freely pursue those ends by looking for the means to them. And then we freely choose those means in order to realize those ends. If this is done well, and if we also have good luck, then we achieve happiness. But if not, then *not*—and yet we are already and always embarked on that rational human fundamental project. For a meaningful and purposeful life is not necessarily a *happy* life, at least in the ordinary sense, or senses, of "happiness."

But what *is* happiness? It seems unexceptionably true, and commonsensical, to say that real human personal happiness is a coherent combination of good experiences, material well-being, salient social status, good work, good play, and good personal relationships. That is what we commonly call "living the good life." If the pursuit of happiness in that sense is done badly, or even if we choose and do things well but are merely unfortunate and unlucky, then we suffer and are unhappy. But an unhappy life is not necessarily a *bad* life. A real human personal life with various ends and means embedded in it, and structuring it, whether it ultimately leads to happiness or to unhappiness, is a life that is *meaningful and purposeful* to that extent.

According to Existential Kantian Ethics, as I have said, morality is the essence of ethics. More specifically, however, according to Existential Kantian Ethics, morality is about what real human persons ought to choose and do (the obligatory), what we ought not to choose and do (the impermissible), and what it is acceptable for us to choose and do even if it is not obligatory (the permissible). What we ought to choose and do is our *duty*.

Hence, according to Existential Kantian Ethics, all morality is inherently "deontological," or concerned with duties.[12]

The moral notion of duty, however, as Existential Kantian Ethics conceives it, needs to be elaborated further in two important ways, in order to avoid two corresponding classical and also commonplace-contemporary misunderstandings of that notion.

First, *duty is not impersonal.* On the contrary, duty as conceived by Existential Kantian Ethics is an intensely personal matter, both on the side of the real human person who is a moral agent and also on the side of those real human persons whose lives are inextricably connected, for better or worse, with that of the moral agent. As W.D. Ross very aptly puts it, in criticizing G.E. Moore's version of Utilitarianism:

> The essential defect of the 'ideal utilitarian' theory is that it ignores, or at least fails to do full justice to, the highly personal character of duty. If the only duty is to produce the maximum of good, the question of who is to have the good—whether it is myself, or my benefactor, or a person to whom I have made a promise to confer that good on him, or a mere fellow man to whom I stand in no such special relation—should make no difference to my having a duty to produce that good. But we are all sure that it makes a vast difference.[13]

Ross is wrong, of course, that *this* is a knock-down criticism of Utilitarianism. It is quite possible to be a Utilitarian and also take various personal, psychological facts to be amongst the utility-maximizing facts.[14] But he is nevertheless absolutely right about what I will call *the human face of duty*, according to Existential Kantian Ethics. So Existential Kantian Ethics is *deontology with a human face*.

Second, and perhaps even more importantly, *duty does not entail any fundamental resistance to human desires.* On the contrary, duty as conceived by Existential Kantian Ethics is the result of choosing and acting on the basis of our deepest and most fundamental desire, the second-order desire for what I call *moral self-transcendence*, namely the desire to be moved by first-order effective desires that are non-hedonistic, non-self-interested, non-selfish, and non-consequentialistic, for the sake of the Categorical Imperative, and for the sake of the absolute, non-denumerably infinite, intrinsic, objective value, or *dignity*, of real persons (see also chapter 2 below). Otherwise put, according to Existential Kantian Ethics, doing one's duty—that is, choosing or acting on the second-order desire for moral self-transcendence—is the same as choosing or acting from the fundamental moral emotion Kant calls "respect" (*Achtung*):

> Duty is the necessity of an action done from respect for the moral law. (*GMM* 4: 400)

Choosing or acting from the second-order desire for moral self-transcendence, or respect, sometimes involves overriding and/or suppressing hedonistic, self-interested, selfish, or consequentialistic first-order desires. But this is only in order to satisfy our

deepest and most fundamental desire, namely the desire for moral self-transcendence. And in any case, it is perfectly possible to do one's duty and *also* satisfy hedonistic, self-interested, selfish, or consequentialistic first-order desires, provided that the following subjunctive conditional or counterfactual is true:

> We would still have done our duty in that actual act-context even if our hedonistic, self-interested, selfish, or consequentialistic first-order desires had *not* been satisfied in that context, by virtue of the second-order desire for moral self-transcendence, that is, by virtue of the fundamental moral emotion of respect.

Or in other words, in that actual context, the second-order desire for moral self-transcendence would have volitionally guaranteed that we structurally mobilized our effective first-order desires in such a way as to do our duty in that context, no matter what our other first-order or higher-order desires were in that context. So doing our duty is perfectly consistent with *our enjoying doing it*, as Ross also very aptly points out:

> The sense of duty tends to be described as the sense that one should do certain acts, though on other grounds (for example, on the ground of their painfulness) one wants not to do them. But "the sense of duty" really means that we ought to do certain acts, whether or not on other grounds we desire to do them, and no matter with what intensity we may desire, on other grounds, not to do them. One of the effects of the forming of a habit of dutiful action is that any natural repugnance one may have to dutiful acts on other grounds trends to diminish. If we form a habit of early rising, for example, it becomes easier, and less unpleasant, to rise early.[15]

The thesis that all morality is deontological, or concerned with duties, should also not be understood in such a way as to exclude moral concern with the *good character* of real human persons (aka "virtues"), or the *good results* of their choices and acts (aka "good consequences"). And in that sense, a deontological ethics can be smoothly compatible with central elements of *virtue ethics* and *consequentialism*. Nevertheless, choice and action can be obligatory, impermissible, and permissible in two sharply different ways:

> either (i) restrictedly, conditionally, and as a means to some other end—which is hypothetical, instrumental obligation,
> or (ii) strictly, unconditionally, and as an end-in-itself—which is categorical, non-instrumental obligation.

The morality of good character, aka virtue ethics, and the morality of good results, aka consequentialism, both fall under the umbrella of *instrumental* obligation, whereby we are required to choose and act certain things as means or tools in order to produce or realize various other good things—for example, virtuous character traits or beneficial

consequences—as ends-in-themselves. By contrast, morality as it is understood by Existential Kantian Ethics takes the obligatoriness, impermissibility, and permissibility of choice and action to be fundamentally and primarily *non*-instrumental, and only derivatively and secondarily instrumental.

That which is an end-in-itself has its goodness or positive value inherently or intrinsically. For example, all those things that can make us happy—enjoying pleasant or otherwise satisfying experiences and pastimes; being healthy; being physically attractive; making lots of money; possessing lots of property, portable or otherwise; being a big fish in an appropriately-sized pond; earning the genuine admiration and respect of others; pursuing and completing significant aesthetic, artistic, or intellectual projects; creating and sustaining companionate, erotic love relationships; creating and sustaining friendships; belonging to a mutually supportive family; having a family of one's own and successfully raising children, and so-on—are ends-in-themselves and have their positive values intrinsically. Such ends-in-themselves can sometimes also be used as means to other ends, however. In other words, they are only relatively and not absolutely ends-in-themselves, because their objective value can also be treated as extrinsic and dependently relational.

But that which is a strict, unconditional end-in-itself has its positive value absolutely intrinsically and objectively. This is the Highest or Supreme Good—the most important thing in the world, and also even, as Kant says, "beyond the world" (*GMM* 4: 393). According to Existential Kantian Ethics, nondenumerably infinite, absolute, intrinsic, objective value exists, and therefore the Highest or Supreme Good also exists, in the form of an innately specified online capacity for what I call *principled authenticity*. Principled authenticity is what Kant himself called "the good will" (*GMM* 4: 393-394), when it is also fully fused with what Kierkegaard called "purity of heart." Otherwise put, it is an essentially embodied good will, that is, a good will that can really and truly be achieved, at least in part and to some degree, only by real human persons living in this thoroughly nonideal natural and social world. Principled authenticity, in turn, to the extent that it is activated and realized even only partially and to some degree, literally embodies and adequately expresses the essence of our rational "human, all too human" nature.

Derek Parfit remarked that "Kant is sometimes thought of as a cold, dry, rationalist. But he is really an emotional extremist."[16] In my opinion, Parfit was absolutely correct about that, although not in precisely the way Parfit intended. In his passion for the moral law within himself—"[t]wo things fill the mind with ever new and increasing admiration and reverence, the more often and more steadily one reflects on them: *the starry heavens above me and the moral law within me*" (*CPrR* 5: 161)—Kant himself sometimes goes too far and becomes a *purist* in the pejorative sense: passionately formalistic and rigoristic, to the point of moral error. But the right corrective for Kant's own emotional extremism, is *not*, as Parfit thought, Henry Sidgwick's instrumental-rational emotional *minimalism*, but instead Kierkegaard's sharply distinct kind of non-instrumental-rational emotional extremism—in three words, *purity of heart*. The corrective fusion of *Kant's* emotional

extremism with *Kierkegaard's* emotional extremism yields the notion of autonomous wholeheartedness, or principled authenticity, and this corrective fusion constitutes, in my opinion, *precisely the right kind of emotional extremism that is needed for morality*.

According to Existential Kantian Ethics, the Highest or Supreme Good is really and truly in the real world, precisely because and just to the extent that rational human animals or real human persons are living in the real world, for better or worse. Now objective values, generally, are in the real world just because what I call "minded animals" are living in the real world. But absolute, nondenumerably infinite, intrinsic, objective values are our specifically real human personal *gift* to the real world, and also—tragically—our *curse* upon the real world, to the extent that we mostly miserably fail to live up to those very values, and very often wickedly choose to ignore them or act contrary to them.

How can I prove all these claims to you, or to a skeptic? My demonstration of these substantive metaphysical theses is neither deductive nor inductive, but instead *ostensive, abductive, transcendental, and neo-rationalistic*. That is, my demonstration is via appeals to moral phenomenology, via appeals to inference-to-the-best-explanation, involving both transcendental arguments and also transcendental explanations, and via rational intuition.[17] It is only by pointing to ourselves as actual living examples, and then by deploying a robust background metaphysical theory of our nature as free and real human persons inherently capable of cognitive and practical agency, that I can make a sound transcendental inference, based on primitive rational insights, to Existential Kantian Ethics's being the best overall explanation of morality. If this transcendental inference is indeed sound, however, then morality is not only about what we have good or contingently sufficient practical reasons to choose and do (hypothetical, instrumental obligation), but more particularly morality is about what we have right or necessarily sufficient—"overriding"—practical reasons to choose and do (categorical, non-instrumental obligation).

Morality according to Existential Kantian Ethics, then, insofar as it is ineluctably embedded, as essential, in the fully meaningful and all-encompassing value-domain of ethics that it inherently governs and structures, is specifically about the highest or supreme values, ideals, and normative standards of real human persons—that is, the highest or supreme ends-in-themselves, and the highest or supreme practical reasons of rational human life. It is also about the obligations and principles of choice and conduct that flow from these highest or supreme values, which thereby in turn constitute a set of low-bar, minimal, or nonideal standards of rational normativity to go along with the corresponding high-bar, maximal, or ideal standards. A real human personal life with the highest or supreme ends-in-themselves, practical reasons, obligations, and principles structurally immanent within it, and also fully incorporated into it, according to The Two-Dimensional Conception of Rational Normativity, is a life that is fully meaningful, and a morally good life.

The very *best* kind of real human personal life would, in turn, be at once fully meaningful, morally good, and also *deeply* happy in that this full meaningfulness and

happiness are related to one another as the jointly constitutive essential form (or immanent structure) and prime matter (or vital stuffing) of one and the same human life. That is what Kant calls "the sole and complete good" (*GMM* 4: 396). The sole and complete good is also *an intersubjective social and political good*, containing a full and rich elaboration of the "Realm of Ends" formulation of the Categorical Imperative (*GMM* 4: 433-436), later also called the "ethical community" in *Religion Within the Boundaries of Mere Reason* (*Rel* 6: 96-100). So in Kant's own philosophy, just as in Existential Kantian Ethics, morality and ethics achieve their ultimate completion in *religion* and *politics*.[18]

But all happiness, whether deep happiness or *non-deep* happiness—that is, happiness that is not immanently structured by morality, whether it is in fact *immoral happiness*, or else just a *shallow happiness* that merely conforms to morality and is only extrinsically related to it—requires *good luck*, and good luck by its very nature is in short supply. Shallow happiness, for example, has moral value, but not *moral worth*. So according to Existential Kantian Ethics, if push comes to shove, then a life including morality and full meaningfulness but also filled with unhappiness due to sheer bad luck still morally exceeds a life of shallow happiness.

Please do not misunderstand me. Of course I think that shallow happiness, in its place, is *perfectly fine and massively preferable to misery*, other things being equal. I like and indeed often crave shallow happiness as much as the next person, and it would be sheer condescension and sanctimoniousness on my part to be too critical of it. Moreover, it would be a truly good thing—although, in certain crucial respects, it would also be inherently limited in its goodness—if any or all of the people in the world who are suffering could instead enjoy lives of even shallow happiness. But at the same time it is true that we ourselves can do substantially *better* than shallow happiness; that we profoundly want *more* than merely shallow happiness not opnly for ourselves but also for those we truly love and for any other persons whose welfare we seriously care about, which should be everyone; and above all that as real human persons, we possess an innately specified capacity to recognize this Highest or Supreme Good (namely, "the good will" in Kant's sense), and also to desire it wholeheartedly (namely, with "purity of heart" in Kierkegaard's sense). This is the innately specified capacity for principled authenticity.

Indeed, as I argued in *Deep Freedom and Real Persons*, chapters 3 and 5, deep happiness is most fully realized in what Kant calls *Selbstzufriedenheit* or *self-fulfillment*—the actual, active, subjectively experienced, and *phenomenologically self-validating* achievement of principled authenticity, at least partially and to some degree (*CPr* 5: 117-119). And this sublime experience remains really possible in a thoroughly unlucky and (in the shallow sense) unhappy life. All the awful, brute actual facts will remain exactly the same; but you can still *radically change your attitude towards those facts*, and thereby *change your life*. I am thinking particularly here of Rilke's cathartic appreciation of the archaic torso of Apollo; of Camus's Sisyphus; and also of what I have called Wittgenstein's "Mystical Compatibilism" in the *Tractatus*.[19] The possibility of existential Kantian self-

fulfillment in the face of sheer bad luck, sharply contrasts with Aristotelian happiness or "flourishing" (*eudaimonia*), which, as Aristotle famously points out in book I of the *Nicomachean Ethics*, necessarily requires good luck. In this sense, in my opinion, Existential Kantian Ethics is much more closely attuned to the realities of real human personal life in a nonideal natural and social world, than Aristotelian ethics is. In chapter 2 below, I will carefully spell out the semi-technical sense in which the actual natural and social world in which we live, move, and have our being, is not merely nonideal, but in fact *thoroughly* nonideal.

1.3 EXISTENTIAL KANTIAN ETHICS VERSUS MORAL RELATIVISM, MORAL SKEPTICISM, AND MORAL PARTICULARISM

Philosophical ethics, aka moral philosophy, has been standardly divided into three parts:[20]

> (i) *meta-ethics*, which deals with metaphysical, semantic, and epistemological issues about morality,
> (ii) *normative ethics*, which deals with different moral theories and moral frameworks, including specific claims about the highest ends and reasons of rational human life, as well as corresponding obligations, prohibitions, and permissions, and
> (iii) *applied or practical ethics*, which deals with normative ethical issues in real-life, real-world contexts.

Now one fundamental triad of questions in meta-ethics is this:

> Are there rationally defensible moral principles?
> Are morally right choice and right action the same as choosing or acting on principle?
> Is the morally best person the person of principle?

Following the terminology of the recent and contemporary debate about these questions, I will call the thesis that rationally defensible moral principles exist, that right choice and right action are the same as choosing or acting on principle, and that the morally best person is the person of principle, *moral generalism*.[21] Building on that, I will call the thesis that not only gives affirmative answers to all three questions, but also asserts that at least some moral principles are absolutely universal and also objective in the sense of holding in every possible set of circumstances and also being intersubjectively accessible for all rational human animals or real human persons, *moral absolutist generalism*.

Moral absolutist generalism is obviously a strong thesis in meta-ethics. But Existential Kantian Ethics counts as an especially strong version of moral absolutist generalism, in

that it not only asserts all of the basic theses of the latter, but also is "existential" in the four senses I sketched in section 1.0. So in those respects, Existential Kantian Ethics is in fact a *super*-strong theory in meta-ethics and normative ethics. More specifically, however, Existential Kantian Ethics makes these seven claims:

> (i) that some rationally defensible moral principles exist,
> (ii) that at least some of the rationally defensible moral principles are absolute and objective in the sense of being both strictly universal and also intersubjectively accessible for all real human persons whatsoever—namely, the set of moral meta-principles collectively labelled the Categorical Imperative,
> (iii) that right choice and right action are the same as choosing or acting on principle,
> (iv) that principled choice or action can be both psychologically and neurobiologically realized in real human persons, at least partially or to some degree,
> (v) that the morally best person is the wholehearted person of principle, namely, the person of *principled authenticity*,
> (vi) corresponding to (iv) and (v), that the life of principled authenticity can be both psychologically and neurobiologically realized in real human persons, at least partially or to some degree, and finally
> (vii) that all real persons, whether human or non-human, are at once the subjects of dignity, where dignity is absolute, nondenumerably infinite, intrinsic, objective value, and also the targets of respect, where *the capacity for respect* is an innately specified capacity for moral emotion that generates the higher-order desire to be moved to choice and action by non-egoistic, non-selfish, non-hedonistic, and non-consequentialist desires and reasons.

It should also be particularly noted that there is a profound sense in which Existential Kantian Ethics is *a morality of human life*—where "life" is defined so as to include both the complex thermodynamic properties of living organisms[22] and also *minded* animal life in the Kantian sense:

> Life is the faculty of a being to act in accordance with the laws of the faculty of desire. The faculty of desire is a being's faculty to be by means of its representations the cause of the reality of the objects of these representations. Pleasure is the representation of the agreement of an object with the subjective conditions of life, i.e., with the faculty of the causality of a representation with respect to the reality of its object (or with respect to the determination of the powers of the subject to action in order to produce the object). (*CPrR* 5: 9)

Thus there is a correspondingly profound sense in which Existential Kantian Ethics is also a *bioethics*. "Bioethics," as it is standardly understood in contemporary philosophy, is a branch of applied ethics that deals with medical ethics, reproductive ethics, or environmental ethics, together with related ethical topics in the life sciences. But in the more profound sense in which Existential Kantian Ethics is a bioethics, over and above and

in addition to its direct application to the familiar topic-domains of medical ethics, reproductive ethics, or environmental ethics, it also bottoms out in ultimate, "meaning-of-life" issues. Therefore the very idea and fact of a rational "human, all too human" life *as it is actually lived by us*—whether understood as organismic and genetic life, as essentially embodied human personal life, as social and political life shared with many real human persons, or as ecological and evolutionary life shared with the other sentient or non-sentient living organisms in a thoroughly nonideal natural and social world, and, correspondingly, the very idea and fact of our own and others' inevitable deaths, our *mortality*—significantly determines and inflects Existential Kantian Ethics.

It should also be particularly noted, before going on, that Kant's *own* ethics, at least as it is presented in the *Groundwork*, as well as the leading recent and contemporary versions of Kantian ethics,[23] are all themselves substantially *less* philosophically ambitious than Existential Kantian Ethics—as presumptuous as that might initially sound. In particular, in the *Groundwork*, Kant is excessively cautious about the metaphysics of free will, claiming only that free will is conceptually or logically possible, and that we necessarily choose and act intentionally under the Idea of our own positive freedom or autonomy (*GMM* 4: 447, 455).[24] To be sure, things are importantly different in the *Critique of Practical Reason*, with its explicit doctrine of "The Fact of Reason," its explicit thesis that natural mechanism is a necessary condition of Natural Determinism—which directly implies that anti-mechanism is a necessary condition of metaphysically real free will, aka *deep freedom*—and, correspondingly, its clear commitment to Source Incompatibilism, namely, the mutual inconsistency of

> (i) metaphysically real free will, which has "*objective* and … undoubted *reality*,"[25]
> and
> (ii) Natural Determinism, at the source of agency (*CPrR* 5: 49).

And things are, yet again, importantly different in the *Critique of the Power of Judgment*, with its clear commitment to *some* fairly robust version of *anti*-mechanism,[26] including both natural teleology and organismic self-production (aka "autopoeisis"), and also spontaneous rational teleology or creative intentionality, as individually necessary and jointly constitutive conditions of practically and transcendentally free will, or autonomy. Indeed, things are even importantly different at certain places in the earlier *Critique of Pure Reason*, where Kant explicitly says that practical freedom can be empirically confirmed (*CPR* A802-803/B830-831). More generally, pretty much everywhere *but* the *Groundwork*, Kant clearly and distinctly presents himself as the defender of a highly original, non-classical version of metaphysical-libertarian incompatibilism.[27]

Nevertheless, the leading recent and contemporary versions of Kantian ethics, taking their cue from the excessive metaphysical caution of the *Groundwork*, and the "hyper-disciplined" character of mainstream Anglo-American professional academic philosophy

in the latter half of the 20th century and well into the second decade of the 21st, are all more or less officially committed to the following three-part thesis, the conjunction of which constitutes contemporary "Soft Determinism":

> (i) freedom and Universal Natural Determinism are mutually consistent (Compatibilism),
> (ii) freedom exists, and
> (iii) Universal Natural Determinism is true.

But in sharp opposition to all that, I think that the commitment to Soft Determinism, although comfortably in line with the professional-philosophical status quo both inside and outside the sub-domain of Kant-scholarship, leaves all these versions of Kantian ethics "in a lonely place."[28]

In the first place, as I just pointed out, everywhere but the *Groundwork*, Kant himself is emphatically *not* a soft determinist.

And in the second place, from my own philosophical point of view, the doctrine of *Natural Libertarianism* that I spell out and defend in *Deep Freedom and Real Persons*, directly entails that real human persons are capable of metaphysically real, robust mental causation. This is the same as to say, using my terminology, that real human persons are "deeply free"; that they are ultimate incompatibilistic sources of autonomous free agency; correspondingly, that their choices and acts have "up-to-me-ness"; and that all of them are, at the very same time, fully embedded in the physical natural world, precisely because autonomous free agency is nothing more and nothing less than a special immanent structural *form* of biological life, hence a life-form that in turn, as inherently self-organizing and purposive, is both explanatorily and ontologically irreducible to any form of natural mechanism, including either Universal Natural Determinism or Indeterminism.

By sharp contrast, as I also argue in *Deep Freedom and Real Persons*, Soft Determinism entails *epiphenomenalism*, or the causal inertness of the mental. Hence for versions of Kantian ethics that also accept Soft Determinism, there can be no such thing as a conscious act, state, or process that *seems to be* autonomous just because it *really and truly is* autonomous. Or more sharply put: for all those leading recent and contemporary versions of Kantian ethics that also accept Soft Determinism, there is no such thing as *metaphysically real, robust* autonomy—autonomy with ultimate incompatibilistic sourcehood, deep freedom, or up-to-me-ness, and all of them fully embedded in the natural or physical world. Instead, there is *only epistemic and psychological, frail* autonomy—"sensitivity to reasons" or "reasons responsiveness," and a correspondingly more or less rich moral psychology—hence at most a naturally mechanistic and in effect *bogus* autonomy. In point of fact, then, for all versions of Kantian ethics that countenance Compatibilism and Soft Determinism, we are really nothing but "biochemical puppets"[29] or "moist robots"[30]—fleshy deterministic or indeterministic Turing machines, enacting Darwinian evolutionary programs—even despite all our "reasons-responsive" bells and

whistles. For all such Kantians, our conscious will is to our animal bodies only as, in Thomas Huxley's stark formulation, a steam whistle is to a steam engine.[31] So instead of deep freedom we have nothing but, to use Kant's equally stark formulation, "the freedom of a turnspit" (*CPrR* 5: 97).

But *how could I ever really and truly be a human person and a moral agent*, if I had only the freedom of a puppet, robot, steam-whistle, or turnspit?[32] Personhood and agency would then be no more than *a dream inside a machine*. Therefore, in my opinion, any version of Kantian ethics that accepts Compatibilism and Soft Determinism is ultimately a philosophical *scandal* in Kant's pregnant sense of that term (*CPR* Bxxxix n.).

This in turn constitutes a fundamental objection to Korsgaard's *Self-Constitution* and Parfit's *On What Matters* alike. To be sure, their moral psychologies, their theories of human agency, their theories of human rationality, and their development of various existential themes centering on integrity and fundamental value, are all deeply insightful and highly thought-provoking. But the shared background metaphysics of their accounts of rational human agency is clearly some or another version of Compatibilism plus Soft Determinism.[33] Therefore, their accounts are both directly subject to compelling counterarguments against any version of either Compatibilism or Soft Determinism, namely, the four arguments for the thesis I call *local incompatibilism with respect to natural mechanism*.[34] So at the end of the day, their accounts are both plain *false*.

Now back to Existential Kantian Ethics, which entails moral absolutist generalism. There are at least three different ways of denying or rejecting this super-strong brand of generalism.

First, there is *moral relativism*,[35] which says

> (i) that there are *no absolutely universal* moral principles,
> (ii) that there are no absolutely universal and objective moral principles precisely because there are no absolutely universal moral principles as such, and
> (iii) that there are and can be only either *individually relative or culturally relative* moral principles, each of which is morally equivalent with all of the others, incommensurable with all of the others, possibly inconsistent with any of the others, and true or false just because that individual or culture believes that it is true or believes that it is false.

Second, there is *moral skepticism*,[36] which says

> (i) that there are *no objective* moral principles or values,
> (ii) that there are no absolutely universal and objective moral principles precisely because there are no objective moral principles or values as such, and
> (iii) that the belief that there are some absolutely universal or objective moral principles or values is nothing but a projection of our conscious or non-conscious desires and wishes, and best explained by evolutionary psychology.

And third, there is *moral particularism*,[37] which says

> (i) that there are *no rationally defensible* moral principles,
> (ii) that there are no absolutely universal and objective moral principles precisely because there are no rationally defensible moral principles as such,
> (iii) that morally right choice and right action *do not consist in applying principles to cases*,
> (iv) that the morally best person is *not the person of principle*,
> (v) that moral reasons vary in content and force from context to context, and also in holistic relationship to collections of contexts,
> (vi) that morally right choice and action consist essentially in particular responses to particular cases in particular contexts, and depend essentially on making good moral judgments in just those cases and contexts, and
> (vii) that even if there are some moral principles and some people who follow them, nevertheless the morally best person is *always the person of good context-sensitive moral judgment, and never the moral rule-monger*.

It is important to note that moral relativism, moral skepticism, and moral particularism are all logically independent of each another. For example, some moral relativists are also moral generalists but not moral skeptics (for example, cultural relativists like the early 20th century Yale sociologist William Graham Sumner[38]); some moral skeptics are not moral particularists (for example, Nietzsche, Mackie, and Richard Joyce, all of whom offer "error-theories" of the construction of supposedly absolutely universal objective moral principles); and some moral particularists are also moral realists but not moral skeptics (for example, Ross, in his views about "actual duties" or "duties proper,"[39] and also contemporary neo-Aristotelians like Jonathan Dancy[40]).

It is also important to note that the properties of universality and objectivity are logically independent—although still mutually consistent—notions. Universality means that a truth or value holds in all (relevant) sets of circumstances, and objectivity means that a truth or value can be known by any rational animal thinker whatsoever, whether human or non-human, and does not depend on any one individual or what she believes or feels. These can obtain together. But some universal truths might be such that they cannot be known by any rational animal thinker whatsoever, either because they are simply unknowable or because knowledge of them depends on some particular indiividuals and what they believe and feel. Contrapositively, a perfectly objective truth might nevertheless also be dependent for its meaning on highly contextual factors (for example, "I am here now"), and so not be universal.

To begin to motivate Existential Kantian Ethics—that is, to show that there are some very good reasons for taking Existential Kantian Ethics and its moral absolutist generalism seriously, even if the philosophical jury is still out as to whether it will endorse Existential Kantian Ethics or not—here are some preliminary critical considerations against moral relativism, moral skepticism, and moral particularism, in turn.

First, here is an argument against moral relativism. Moral relativism divides into two basic kinds:

(i) individual relativism, and
(ii) cultural, aka communitarian, relativism.

The thesis of individual relativism says that there is no such thing as universal objective moral truth and that whatever an individual believes is morally right or wrong, truly is morally right or wrong. By contrast to individual relativism, the thesis of cultural or communitarian relativism says that there are no such things as absolutely universal objective moral principles, that none of the many different culturally (or community-) specific moral codes (including ours) has any special moral status because all of them are morally equivalent, that each culture's or community's moral codes are incommensurable with all the others and can be inconsistent with any of the others, and that the moral beliefs of each culture or community strictly determine what is morally right for that culture or community, and/or morally wrong for that culture or community. Correspondingly, the primary argument for the truth of moral relativism, whether individual or cultural/communitarian, runs as follows:

(1) As a matter of empirical fact, different people have different and often conflicting moral beliefs about moral principles, and different cultures or communities have different and sometimes conflicting moral beliefs about moral principles.
(2) Therefore there are no absolutely universal and objective moral principles.
(3) Therefore there are and can be only either individually relative or culturally/community- relative moral principles, each of which is morally equivalent with all of the others, incommensurable with all of the others, possibly inconsistent with any of the others, and true or false just because that individual or culture/community believes that it is true or false.
(4) Therefore, moral relativism is true.

Step (1) is of course true. But obviously (2) does not follow as a logical consequence from (1). It is quite true that from the fact that two or more different beliefs about *X* are mutually logically inconsistent, it does indeed follow that at least one of the beliefs must be false, because they cannot all be true. Nevertheless, from the fact of two or more different beliefs about X, some of which are mutually logically inconsistent, precisely nothing follows about the nature of X. Belief in proposition P does not itself entail the truth of P: the fact that P is not entailed by the mere fact of someone's belief that P, nor is it entailed by the mere fact that a great many people believe that P. Correspondingly, a belief in the denial of proposition P does not itself entail the falsity of P: the fact that not-P is not entailed by the mere fact of someone's belief that not-P, nor is it entailed by the mere fact that a great

many people believe that not-P. So even given the truth of (1), there could still be absolutely universal objective moral principles.

Moreover, even if there were no absolutely universal objective moral principles, it would not follow that there are only individually relative or culturally/community- relative moral principles. That is because even if there were no absolutely universal objective moral principles, there could still be objective moral principles

> (i) that held for *a great many* contexts, individuals, and cultures or communities, even if not strictly speaking holding for absolutely all contexts, individuals, and cultures or communities, and
> (ii) that in addition could hold for *any* context, individual, or culture/community whatsoever, provided that certain favorable background conditions obtain, and hence they would hold "other things being equal," or ceteris paribus.

Here we can think, for example, of the moral principles "It is impermissible to kill innocent people" and "It is impermissible to tell lies constantly." There are obviously contexts in which these might not hold—for example, killing one innocent person in order to save five innocent others, as in The Trolley Problem (which will be discussed in detail in chapter 5 below), or lying constantly when operating as an undercover agent in a morally warranted war against an evil enemy. But on the whole, and other things being equal, these moral principles certainly do seem to be objectively valid: that is, they certainly do seem to hold for a great many contexts, individuals, and cultures or communities. I will call such non-absolutely universal, yet still objectively valid, and genuinely although restrictedly universal—under ceteris paribus conditions—moral principles, *fairly universal and objective moral principles*. Given the possibility of fairly universal and objective moral principles, the truth of moral relativism obviously does not follow from (1) and (2).

Furthermore, there is a classical and obvious problem with both individual and also cultural or communitarian moral relativism, having to do with truth and logical consistency. If relativism were true, then if person A or culture/community C1 believes that principle P is true, then P is true. But if person B or culture C2 also believes that principle P is false, then P is false. So according to individual or cultural/communitarian relativism, principle P could be both true and false. Indeed, according to individual or cultural/communitarian relativism, every moral principle could be both true and false. But that is absurd and unintelligible, precisely because it violates the logical Principle of Non-Contradiction in its logically thinnest and at the same time absolutely unrevisable version,

> Minimal Non-Contradiction:
>
> Accept as truths in any language or logical system only those statements which do not entail that it and all other statements in any or all languages or logical systems are both true and false.[41]

So moral relativism is false.

Nowadays, it is a standard response of the relativist to claim that *for them*, the words 'true' and 'false' mean *true for them* and *false for them*, and not anything objective. But it would not even be correct for the individual relativist or cultural/communitarian relativist to hold that a given moral principle is true or false for them, *just in virtue of the fact that they believe it*. This is because *belief does not, in and of itself, itself entail truth*, whether truth about the larger world, truth about oneself, or truth about one's own culture/community. All that can be validly concluded from the fact that a given individual or culture/community believes a certain principle, is that this individual or this culture/community *indeed believes this principle*. It does not follow that the principle itself is true, whether for them or anyone else. So the thesis

> S believes that P, but since P is other than just the claim "S believes that P," then P can be false

is a logically or conceptually necessary truth about the notions of belief and truth.

Now the next move in the contemporary debate about moral relativism is for the relativist to try to define the meaning of 'truth' in terms of individual or cultural/communitarian belief—say, as *warranted assertibility*, or whatever. But any such move, although it may suffice for playing interesting dialectical games in professional academic philosophy, runs directly into the self-evident contrary fact that any version of the argument

> (1) X believes that P and P is not just the claim "X believes that P."
> (2) Therefore, P is true.

is a fallacious argument. Or in other words, on no even *remotely* plausible construal of the meanings of 'believes' and 'true', does *believes* logically or conceptually entail *true*.

Second, here is an argument against moral skepticism. Moral skepticism challenges the very idea of absolutely universal and objective moral principles and values. One way of being a moral skeptic would be to deny that there are any absolutely universal moral principles or values. Another way of being a moral skeptic would be to deny that there are any objective moral principles or values. This is precisely what Mackie does in his well-known book, *Ethics: Inventing Right and Wrong*. Of course it also follows from the rejection of all objective moral principles or values, that there are no absolutely universal and objective moral principles. But notice that it would *not* follow from the assertion that *there are no absolutely universal and objective moral principles or values*, that there are no objective moral principles. As we saw above, there could still be *some fairly universal*

and objective moral principles, and of course, correspondingly, there could also still be *some fairly universal and objective moral values*.

Now Mackie offers three basic reasons for denying the objectivity of all moral beliefs, claims, or judgments.

First, he appeals to the fact of individual and cultural/communitarian differences in moral beliefs/claims/judgments, and how they contradict one another. This of course is similar to the first two steps of the classical argument for moral relativism.

Second, he says that if there really were objective moral truths or values, then they would have to be metaphysically and epistemologically "queer" because they would not be natural facts, hence they would be neither knowable in ordinary ways through sense perception or the sciences, nor explicable in the ways we normally do in the sciences.

And third, he says that the commonsense belief in objective moral truths or values can be explained away by appealing to an "error theory" of how such a belief came to be. An error theory of the belief in some supposed fact *X* says

> not only (i) that *X* is actually bogus and a myth,
> but also (ii) that we can offer a scientifically acceptable explanation—via, for example, evolutionary psychology—of how a belief in the myth of *X* came to be widely held.

The basic idea here is that we can use the psychological fact of "unconscious projection" together with , for example, a theory about evolutionary mechanisms, in order to explain how our own desires, needs, and wishes to have objective moral truths and values, have unconsciously led us to project them onto the world, even though they do not actually exist there.

Is Mackie's version of moral skepticism correct? On the one hand, he has formulated his claims very carefully, so it is not at all easy to find decisive, simple objections to his basic reasons for skepticism. Refuting the "queerness" argument would require an all-out critique of *scientific naturalism*; and refuting the "error theory" argument would require an all-out critique of *philosophical debunking strategies*. There is no a priori reason whatsoever to think that both critiques could *not* be successfully carried out, but they would of necessity be fairly long-winded and rationally strenuous. On the other hand, however, one way of decisively and simply demonstrating that Mackie is incorrect would be to provide a crisp, compelling, positive argument for the existence of at least *some* absolutely universal and objective moral principles and values—that is, to provide a crisp, compelling, positive argument for moral absolutist generalism. So here is my attempt at that.

An Argument for Moral Absolutist Generalism

(1) There are at least some moral rules that every individual and every culture or community whatsoever must believe and hold in common, at least implicitly, hence even

if not self-consciously. I will call such rules strictly common moral rules. One such strictly common moral rule is this one, which I'll call The Platinum Rule:

"It is impermissible to kill any or all arbitrarily-chosen innocent people for no good reason whatsoever."

Another such strictly common moral rule is this one, which I'll call *The Platinum-Plus Rule*:

"It is impermissible to treat any or all arbitrarily-chosen innocent people either as mere means to others' ends or as mere things, like garbage or offal, for no good reason whatsoever."

(2) That these moral rules—namely, The Platinum Rule and The Platinum-Plus Rule— are indeed strictly common is shown by the following. Any individual who, or any culture or community that, attempted to disbelieve or disobey one of these moral rules would also have to believe that it is morally permissible to kill herself or anyone or everyone else in their own culture, even though they are innocent, as the result of an arbitrary choice, for no good reason at all, and also that it is morally permissible to treat herself or anyone or everyone else in their own culture/community, even though they are innocent, either as mere means or as mere things, like garbage or offal, as the result of an arbitrary choice, for no good reason at all. But, clearly and distinctly, the very idea of the *innocence* of a person entails, at the very least, that she does not morally deserve in any way to be killed or treated either as mere means or as mere things, like garbage or offal, as the result of an arbitrary choice, for no good reason whatsoever. So violating The Platinum Rule and The Platinum-Plus Rule would imply that it is morally permissible to treat oneself or anyone or everyone else, insofar as they are innnocent people, in ways that are not morally deserved in any way, or on the basis of any good reasons whatsoever, as the result of an arbitrary choice. But that is absurd and the moral equivalent of "1=0." Therefore The Platinum Rule and The Platinum-Plus Rule must be believed or obeyed, at least implicitly and even if not self-consciously, by every individual and every culture whatsoever. So they are strictly common moral rules. The moral and rational force of these strictly common moral rules should be directly compared to the logical force of Minimal Non-Contradiction.
(3) The best overall explanation for these strictly common moral rules is that they express absolutely universal and objective moral principles and values.
(4) Therefore there are at least some absolutely universal and objective moral principles and values.
(5) Therefore moral absolutist generalism is true and moral skepticism is false.

Third and finally, here is an argument against moral particularism. The primary argument for moral particularism runs as follows:

(1) Essentially the same set of normative facts can give rise to different sufficient moral reasons for moral judgment in different contexts (aka "the variability of reasons").

(2) Moral reasons for moral judgment are inherently context-sensitive and the specific character of their normative content and action-guiding force is solely and wholly determined by how contexts relate to one another (aka "the holism of reasons").
(3) People who always try to follow the same moral principles in every context are moral rule-mongers and apt to make bad moral decisions.
(4) Therefore, moral particularism is true.

Let us suppose that steps (1) to (3) are all true. Still, (4) does not follow from them, because moral particularism is false. And here are two basic reasons for asserting the falsity of moral particularism.

First, if moral particularism were true, then it would follow that principled consistency in moral choice and action is not an essential, or even good-making, feature of moral rationality. Now particularists are entirely correct in making the point that firmness and unshakeability in a person's character and conduct are not the same as principled consistency, and can hold independently of the latter. But that point is a double-edged sword. This is because it entails that firmness and unshakeability in a person's character and conduct are as context-sensitive and context-determined as any other features of a morally good judge. Moreover, while it is true that firmness and unshakeability in character are not the same as principled consistency, it is also true that one of the several ways in which a person's character can be a contrary opposite of firm and unshakeable is for it to be completely feckless, flaky, and inconsistent. Therefore, according to moral particularism, even a person who was completely feckless, flaky, and inconsistent in her choices and actions from context to context could *still* be a morally good judge. But that is absurd. So by reductio, moral particularism is false.

Second, defenders of moral particularism apparently cannot tell us why the particular moral reasons supporting good moral judgments are themselves *good* particular moral reasons, without either arbitrarily stopping the regress of reasons prior to justificatory bedrock or else implicitly appealing to some or all of the very moral principles—that is, inherently general, cross-contextual moral reasons—they purport to be rejecting. So moral particularism either fails to justify its moral reasons or else it presupposes moral generalism. So either way, moral particularism is rationally unacceptable.

To be sure, defenders of moral relativism, moral skepticism, and moral particularism will not give up their positions easily,[42] and will therefore want to provide various sorts of replies to my critical arguments.[43] So my criticisms are certainly not intended to be *decisive* refutations of those doctrines—on the contrary, as I said above, they are intended only to be preliminary critical considerations in support of Existential Kantian Ethics and its commitment to moral absolutist generalism, and therefore sufficiently good reasons for taking Existential Kantian Ethics seriously, even if the philosophical jury is still out.

1.4 EXISTENTIAL KANTIAN ETHICS VERSUS EGOISM AND ACT CONSEQUENTIALISM

Two other serious philosophical opponents of Existential Kantian Ethics are ethical egoism and act consequentialism.[44] Ethical egoism says that the highest human good is individual self-interest, and that one ought always to choose and act in such a way as to promote one's own self-interest. Act consequentialism says that the highest human good is good results, and that one ought always to choose and act in such a way as to bring about good results—and more specifically, good results in terms of (shallow) happiness for as many people as possible, if you are also a Utilitarian.

The whole of this book, in effect, is an extended defense of Existential Kantian Ethics against ethical egoism and act consequentialism. In carrying out this extended defense I am going to ask, and then attempt to give, intelligible and defensible answers to the following hard questions:

> What are the absolutely universal and objective moral principles?
> How are these absolutely universal and objective moral principles possible?
> How do these absolutely universal and objective moral principles inherently guide right choice and right action?, and
> Why is the wholehearted person of principle the morally best person?

These answers, in turn, will entail my making various substantive claims in meta-ethics, normative ethics, and applied ethics. In order to defend these substantive claims, I will assume, draw directly upon, and also explicitly deploy both the metaphysics of free agency and persons that I worked out and defended in *Deep Freedom and Real Persons*,[45] and also the theory of mental content, cognition, and knowledge I worked out and defended in *Cognition, Content, and the A Priori*.[46]

More precisely, however, I am going to argue that the two co-essential conceptual keys to Existential Kantian Ethics are

> (i) the "skinny" logic and "fat" semantics of moral principles in a thoroughly nonideal natural and social world, and
> (ii) a certain core set of existential Kantian insights about the meaning of a rational human life.

Or otherwise put, I am saying that just as theoretical rationality is grounded in categorically normative logical principles, so too practical rationality is grounded in categorically normative moral principles, and that the two sets of absolute principles are ultimately the same set of absolute principles, also known collectively as the Categorical Imperative, insofar as it is realized in our "human, all too human" rational minded animal lives.

Furthermore, the meaning of life for a rational human minded animal or real human person lies in her lifelong pursuit of principled authenticity in a fully finite, ineluctably contingent, intensely exciting, overwhelmingly lovely, and sometimes also heart-crushingly confusing, contradictory, dangerous, evil, ugly, and tragic world, together with all the other rational animals or real persons, as members of a single universal intersubjective ethical community—all of us individually and collectively committed to mutual respect, mutual equal consideration, mutual aid, and mutual kindness—along with all the non-rational, human or non-human, sentient or non-sentient living organisms, considered now as "associate members" of that universal ethical community. Principled authenticity, in turn, is nothing more and nothing less than a rational human minded animal's or real human person's at least partial or to-some-degree wholehearted autonomous adherence to a special set of guiding principles, all subsumed under the Categorical Imperative, a set of principles that defines her unified life-project in this thoroughly nonideal natural and social world of ours, over the course of her complete, finite, and unique life.

Two very important critical conclusions follow directly from the basic theses of Existential Kantian Ethics.

First, ethical egoism is false because at least some of the actual or possible choices and acts morally required of the wholehearted person of principle are inconsistent with her own self-interest. For example, if faced with a compelled choice between

> either (i) a guaranteed life of shallow happiness for herself, at the cost of her treating someone else either as mere means to others' ends or as a mere thing, that is, like the Nazis treated people, provided that no one else ever finds out,
> or (ii) a guaranteed life of pointless suffering for herself, just because she refuses to treat someone else as a mere means or a mere thing, like the Nazis treated people, even if it were guaranteed that no one else would ever find out that she had treated someone that way,

then the wholehearted person of principle will choose option (ii) and take the awful hit. Of course, option (i) is the ethical egoist option, and option (ii) is the anti-egoist option.

Second, act consequentialism is also false because at least some of the actual or possible choices and acts morally required of the wholehearted person of principle do not bring about good results—and more specifically, they do not bring about good results in terms of shallow happiness for as many people as possible. For example, if faced with a compelled choice between

> either (i) guaranteed lives of shallow happiness for almost everyone else, at the cost of her treating someone else as a mere means or a mere thing, like the Nazis treated people, provided that no one else would ever find out,
> or (ii) exactly the same sort of actual lives for everyone else (that is, no shallowly happier lives for anyone else) plus a guaranteed life of pointless suffering for herself, just because

> she refuses to treat someone else as a mere means or a mere thing, like the Nazis treated people, even if no one else would ever find out that she had treated someone that way,

then the wholehearted person of principle will again choose option (ii) and take the awful hit. Of course, option (i) is the act consequentialist option and option (ii) is the anti-act consequentialist option.

It should be particularly noted that, in order to show that act consequentialism is false, the choice does *not* have to be between

> either (i) guaranteed lives of shallow happiness for almost everyone else, at the cost of her treating someone else as a mere means or a mere thing, like the Nazis treated people, provided that no one else would ever find out,
> or (ii*) *guaranteed lives of pointless suffering for everyone else* plus a guaranteed life of pointless suffering for herself, just because she refuses to treat someone else as a mere means or a mere thing, like the Nazis treated people, even if no one else would ever find out that she had treated someone that way.

This is because (i) is *already* clearly obligatory according to act consequentialism. Nevertheless, it is true that if she were faced with a choice between (i) and (ii*), then the wholehearted person of principle would *still* choose (ii*). And this may seem initially shocking.

But the initial shock is significantly mitigated when one realizes that a life of pointless suffering can be converted *at any time* into a life of principled authenticity, by means of a free choice of the higher-level or Kantian rational human animal. That is the profound lesson of Sisyphus, as interpreted by Camus, and also of Wittgenstein's Mystical Compatibilism; and I also argue this explicitly in chapter 4 below. Therefore in choosing as she does, even though by hypothesis it leads to a guaranteed life of pointless suffering for everyone, the person of principle is not *intentionally* inflicting a life of pointless suffering *on anyone*. If it were in any way really possible for her, then she herself would take on the guaranteed life of pointless suffering alone, and spare everyone else that fate, as the horribly unjust punishment for her refusal to treat someone else either as a mere means or as a mere thing. So the classical "doctrine of the double effect" can be legitimately invoked in this context.

1.5 "THE WORLD OF THE HAPPY": EXISTENTIAL KANTIAN ETHICS VERSUS ETHICAL NATURALISM[47]

> In the beginning was the Act.[48]

> If good or bad willing changes the world, it can only change the limits of the world, not the facts; not the things that can be expressed in language. In brief, the world must thereby become quite another. It must so to speak wax or wane as a whole. The world of the happy is quite another than that of the unhappy.[49]

The meta-ethical doctrine of *ethical naturalism* says that moral facts supervene on natural facts; and ethical naturalism, whether reductive or non-reductive, is the standard view in mainstream contemporary moral philosophy.[50] But, in light of Existential Kantian Ethics, ethical naturalism can be decisively demonstrated to be *false*, by reformulating and updating G.E. Moore's classical "Open Question Argument" against instances of "the naturalistic fallacy."

Let me now explain this claim in more detail. The locus classicus of Moore's Open Question Argument against the naturalistic fallacy is *Principia Ethica*, and his general target is what he explicitly calls "naturalism" in ethics:

> [Naturalism] consists in substituting for "good" some one property of a natural object or of a collection of natural objects; and in thus replacing Ethics by some one of the natural sciences. In general, the science thus substituted is one of the sciences specially concerned with man.... In general, Psychology has been the science substituted, as by J. S. Mill. [51]

And his argument centers on the naturalistic fallacy, defined as follows:

> [T]he naturalistic fallacy ... [is] the fallacy which consists in identifying the simple notion which we mean by "good" with some other notion.[52]

> [The naturalistic] fallacy, I explained, consists in the contention that good *means* nothing but some simple or complex notion, that can defined in terms of natural qualities.[53]

In other words, according to Moore, ethical naturalism is the claim that the property[54] of being good is identical with some simple or complex natural property (which for our purposes we can construe as either a first-order physical property, a second-order physical property, or a sensory experiential property); and the naturalistic fallacy consists precisely in accepting such an identification of properties.

So far, so good—awful pun fully intended. But now for the sad part of the philosophical story.

Many post-Moorean analytic philosophers have accepted Moore's characterization of ethical naturalism, and many have also accepted his anti-naturalistic conclusions. Yet his main argument in support of its putative fallaciousness—The Open Question Argument — is generally held to be a notorious failure. Here is The Argument in Moore's own words:

> The hypothesis that disagreement about the meaning of good is disagreement with regard to the correct analysis of a given whole, may be most plainly seen to be incorrect by consideration of the fact that, whatever definition be offered, it may always be asked, with significance, of the complex so defined, whether it is itself good.[55]

> We must not, therefore, be frightened by the assertion that a thing is natural into the admission that it is good: good does not, by definition, mean anything that is natural; and it is always an open question whether what is natural is good.[56]

For convenience, I will call the fundamental ethical property of being good, "The Good." The Open Question Argumen then says that any attempt to explain The Good solely in terms of some corresponding natural property N (say, the property of being a pleasurable state of mind), automatically falls prey to the decisive objection that even if X is an instance of N it can still be significantly asked *whether* X is good: that is, it can be significantly postulated that X is an instance of N but X is *not* good. So The Good is not the same as N. Moore's rationale for this is that the only case in which it would be altogether nonsensical to postulate that X is an instance of N but X is not good, is the case in which it is strictly impossible or contradictory to hold that X is not good, that is, when X is, precisely, good. So if it is significant to ask whether X is N but not good, then N is not identical to The Good. And Moore finds it to be invariably the case that it is significant to ask whether X is N but not good, hence invariably the case that N is not identical to The Good. He concludes that The Good is an indefinable or unanalyzable non-natural property, and that it is a fallacy to try to identify The Good with any natural property.

In my opinion, Moore's classical Open Question Argument is doomed because of a mistake he has made about the individuation of properties. The problem, as I see it, is that the argument implies a criterion of property-identity that is absurdly strict.[57] Familiar criteria for the identity of two properties include

(i) necessary equivalence of their analytic definitions,
(ii) synonymy of their corresponding predicates, and
(iii) identity of their cross-possible-worlds extensions.

But Moore's criterion is importantly different:

> [W]hoever will attentively consider with himself what is actually before his mind when he asks the question "Is pleasure (or whatever it may be) after all good?" can easily

> satisfy himself that he is not merely wondering whether pleasure is pleasant. And if he will try this experiment with each suggested definition in succession, he may become expert enough to recognise that in every case he has before his mind a unique object, with regard to the connection of which with any other object, a distinct question can be asked. Everyone does in fact understand the question "Is this good?" When he thinks of it, his state of mind is different from what it would be, were he asked "Is this pleasant, or desired, or approved?" It has a distinct meaning for him, even though he may not recognize in what respect it is distinct. Whenever he thinks of "intrinsic value," or "intrinsic worth," or says that a thing "ought to exist," he has before his mind the unique object--the unique property of things--which I mean by "good".... "Good," then is indefinable.[58]

In other words, Moore's criterion is that two properties are identical if and only if the intentional contents of the states of mind in which the properties are recognized, are phenomenologically indistinguishable.[59] Consequently, even two properties that are by hypothesis definitionally equivalent—for example, the property of being a bachelor, and the property of being an adult unmarried male—will come out non-identical according to this test. The intentional content of the state of mind of someone who says or thinks that *X* is a bachelor is clearly phenomenologically distinguishable from that of the same person when she says or thinks that *X* is an adult unmarried male. I might not wonder even for a split second whether a bachelor is a bachelor, yet find myself mentally double-clutching as to whether a bachelor is an unmarried adult male. But then according to that test it is not nonsensical to ask whether *X* is an unmarried adult male but not a bachelor: from which we must conclude by Moorean reasoning that the property of being a bachelor is indefinable, and that it is a fallacy to try to identify any property with any other property, *including the property that expresses its definition*. Obviously this cannot be correct. It is patently absurd to constrain property identity so very, very tightly.[60]

Moore's ethical anti-naturalism also contains another less noticed but equally serious difficulty. This difficulty stems from his explicit commitment to a certain strict modal connection between intrinsic-value properties and natural facts:

> I have tried to shew, and I think it is too evident to be disputed, that such appreciation [of intrinsically valuable, or good, qualities] is an organic unity, a complex whole; and that, in its most undoubted instances, part of what is included in this whole is *a cognition of material qualities*, and particularly of a vast variety of what are called *secondary* qualities. If, then, it is *this* whole, which we know to be good, and not another thing, then we know that material qualities, even though they be perfectly worthless in themselves, are yet essential constituents of what is far from worthless.... [A] world, from which material qualities were wholly banished, would be a world which lacked many, if not all, of those things, which we know most certainly to be great goods.[61]

> [I]f a given thing possesses any kind of intrinsic value in a certain degree, then not only must that same thing possess it, under all circumstances, in the same degree, but also

> anything *exactly like it*, must, under all circumstances, possess it in exactly the same degree. Or, to put it in the corresponding negative form: it is not *possible* that of two exactly similar things one should possess it and the other not, or that one should possess it in one degree, and the other in a different one.[62]

According to Moore, then,

> (i) every intrinsic-value property has some complex set of natural qualities as its "essential constituents," and
> (ii) for any natural thing that "possesses any kind of intrinsic value in a certain degree, then not only must that same thing possess it, under all [logically possible] circumstances, in the same degree, but also anything *exactly like it*, must, under all [logically possible] circumstances, possess it in exactly the same degree."

So, in effect, according to Moore, intrinsic-value properties are both constituted by and also logically strongly supervenient on natural properties. It follows that The Good is, incoherently, both natural and also non-natural. I say "incoherently" rather than "inconsistently" because, strictly speaking, it is possible to hold that two sets of properties are non-identical even though one of those sets of properties is logically strongly supervenient on the other set of properties. But since logical strong supervenience implies both explanatory reduction and also ontological reduction, even if not strict identity, and since the philosophical upshot of Moore's ethical anti-naturalism is surely intended to be not the mere non-identity of The Good with any other property, but rather the explanatory and ontological irreducibility of The Good to any other property, then his overall view is in conflict with itself.

We have just seen that Moore's ethical anti-naturalism is a double failure. But all is not lost, for this double failure teaches us two important philosophical lessons.

First lesson: Do not make your argument against ethical naturalism rest on questionable assumptions about property-individuation or property-identity.

Second lesson: You must directly attack ethical naturalism's strong supervenience thesis.

Taking these two post-Moorean dicta to heart, here is a new general argument against ethical naturalism, grounded on Existential Kantian Ethics.

1.5.1 The First Naturalistic Fallacy as Failed Logical Supervenience

For the purposes of my argument I will need only four basic assumptions.

First, I will need the familiar metaphysics of *strong supervenience*, briefly characterized and defined below.

Second, I will need the equally familiar "conceivability entails possibility" non-supervenience argument strategy deployed by David Chalmers and many others in the context of recent philosophy of mind, but assuming only the truth of my own modal semantic framework, based on a positive theory of the analytic-synthetic distinction,[63] and not the truth of Chalmers's "Two Dimensional" modal semantics, which, following George Bealer, I regard as highly questionable.[64]

Third, I will need the intrinsically compelling, basic authoritative Kantian moral rational intuition that arbitrarily torturing completely innocent people to death, like the Nazis did, for no good reason whatsoever, is a direct violation of the second formulation of the Categorical Imperative, namely, The Formula of Humanity as End-in-Itself,

> 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. (*GMM* 4: 429),

and, as such, it is self-evidently morally wrong.

Fourth and finally, I will need the following thesis about the nature of basic intentional action:

> *A* is a basic intentional act of an essentially embodied human person *P* if and only if *A* is an intentional body movement of *P* that is structurally caused and actively guided and controlled by *P*'s simultaneously *trying* to perform *A*, which in turn is a physically irreducible conscious effective first-order desire to perform *A*, which in turn is *P*'s *will*.

Michelle Maiese and I have argued at length and in detail for this thesis in *Embodied Minds in Action*.[65] To put the thesis in a name-dropping context, however, it is a constructive extension of Harry Frankfurt's *hierarchical desire* conception of the will[66] and also Frankfurt's *guidance-control* conception of intentional action,[67] together with Brian O'Shaughnessy's action-theoretic notion of *trying*,[68] framed against the backdrop of The Essential Embodiment Theory of the mind-body relation.

Granting those assumptions, then the first ethical-naturalist claim I am putting forward for refutation is this one:

> The right (namely, the "ought") logically supervenes (globally or regionally or locally) on natural facts (namely, the "is").

Now this claim, I will argue, is false. But to believe that the logical supervenience of the right and the ought on natural facts is *true*, given that it is actually *false*, is what I will call *the first naturalistic fallacy*, by way of clarifying and precisifying one important version of Moore's basic idea.

The thesis of logical supervenience says that the existence and specific character of *B*-facts logically supervene on *A*-facts. Now *B*-facts logically supervene on *A*-facts if and only if

> (i) A-facts logically necessitate B-facts,
> (ii) B-facts are either downwards identical to A-facts, or not downwards identical to A-facts, yet
> (iii) logically necessarily, there can be no change in any of X's B-properties without a corresponding change in X's A-properties, and
> (iv) logically necessarily, any two beings that are A-property indiscriminable are also B-property indiscriminable (but not necessarily conversely—in case B-facts are not downwards identical to A-facts).

The domain of *A*-facts is *the supervenience base* and the domain of *B*-facts is *the supervening domain*. And here are the main implications of a logical supervenience thesis: Fix all the *A*-facts and then you have thereby fixed, with a priori logical necessity (that is, non-empirically holding in every logically possible world), all the *B*-facts. Or otherwise put: Know everything there is to know about the *A*-facts, and you thereby know, conceptually a priori, everything there is to know about the *B*-facts.

I would now like to reformulate and update Moore's classical Open Question Argument as an "analytic conceivability entails logical possibility" non-supervenience argument against *reductive* ethical naturalism.

But first I will define reductive ethical naturalism. Reductive ethical naturalism is the disjunction of reductive consequentialism (whether act or rule consequentialism), reductive hedonism, reductive ethical egoism, and something I will rather inelegantly call *reductive ethical hybridism*—

> *Reductive consequentialism*: All the moral facts are nothing but good-consequence facts.
> *Reductive hedonism*: All the moral facts are nothing but positive-pleasure facts.
> *Reductive ethical egoism*: All the moral facts are nothing but in-my-best-self-interest facts.
> *Reductive ethical hybridism*: All the moral facts are nothing but some hybrid mixture of good-consequence facts, positive-pleasure facts, or in-my-best-self-interest facts.

Assuming that disjunctive definition, then here is the basic form of my Existential Kantian Ethics-grounded, reformulated, and updated classical Moorean Open Question Argument against reductive ethical naturalism:

> (1) Purportedly, the domain of *B*-facts (= the right) logically supervenes on the domain of *A*-facts (= some or another set of natural facts).
> (2) But analytically conceivably, hence logically possibly, all the same *A*-facts (= the same set of natural facts) can obtain in another minimally physically duplicated possible world, but not all the same *B*-facts obtain in that world (= not right in that world)?

> (3) If yes, then *B*-facts (= the right) do not logically supervene on *A*-facts (= some or another set of natural facts), hence the thesis that moral facts (= the right) logically supervene on some or another set of natural facts commits the first naturalistic fallacy and is false.

Correspondingly, here is my Existential Kantian Ethics-grounded, reformulated, and updated Open Question Argument, itself, as an "analytic conceivability entails logical possibility" non-supervenience argument against reductive ethical naturalism:

> (1) Purportedly, the domain of *B*-facts (= the right) logically supervenes on the domain of *A*-facts (= either good-consequence facts or positive-pleasure facts or in-my-best-self-interest facts, or some hybrid mixture of them (= some or another set of natural facts).
> (2) But analytically conceivably, hence logically possibly, all the same good-consequence facts or positive-pleasure facts or in-my-best-self-interest facts, or some hybrid mixture of them (= the same set of natural facts) can obtain in another minimally physically duplicated logically possible world, but not all the same *B*-facts obtain in that world (= not right in that world)?
> (3) If yes, then *B*-facts (= the right) do not logically supervene on *A*-facts (= good-consequence facts or positive-pleasure facts or in-my-best-self-interest facts, or some hybrid mixture of them (= some or another set of natural facts), hence the thesis that moral facts (= the right) logically supervene on some or another set of natural facts (= good-consequence facts or positive-pleasure facts or in-my-best-self-interest facts, or some hybrid mixture of them) commits the first naturalistic fallacy and is false.

For example: Pick any basic act *A* in the actual world and fix any set of good-consequence facts or positive-pleasure facts or in-my-best-self-interest facts, or any hybrid mixture of them, relative to *A*, and also assume that *A* is prima facie right in the actual world just by virtue of fixing those facts, and also assume that the specific character of *A* is whatever it adventitiously happens to be. Nevertheless, it is still analytically conceivable and therefore logically possible that *A* is morally wrong in some *logically possible minimal* good-consequence facts or positive-pleasure facts or in-my-best-self-interest facts or some-hybrid-mixture-of-them facts *physical duplicate* of the actual world in which the specific character of *A* spontaneously varies. To show this, let us say that *A* is now specifically an attempt arbitrarily to torture some completely innocent person *P* to death, like the Nazis did, for no good reason at all. Even despite being prima facie right in the actual world on consequentialist, hedonistic, egoistic, or hybrid grounds, *A* is always wrong, no matter what. So neither good-consequence facts nor positive-pleasure facts nor in-my-best-self-interest facts, nor any hybrid mixture of them, constitutes the determining ground of rightness, and reductive ethical naturalism is false. The determining ground of moral rightness is nothing more and nothing less than the essentially embodied basic intentional act of a rational animal, for better or worse, insofar as it is immanently structured by the Categorical Imperative.

1.5.2 The Second Naturalistic Fallacy as Failed Nomological Supervenience

The second ethical-naturalist claim I am putting forward for refutation is this one:

> The right (namely, the "ought") nomologically supervenes (globally or regionally or locally) on natural facts (namely, the "is").

Now this claim too, I will argue, is false. But to believe that the nomological supervenience of the right and the ought on natural facts is *true*, given that it is actually *false*, is what I will call *the second naturalistic fallacy*, by way of clarifying and precisifying a *second* important version of Moore's basic idea.

Having previously extended the notion of strong supervenience to logical supervenience, I want now to extend that notion to *nomological* strong supervenience, or nomological supervenience for short. Nomological supervenience is the thesis that the existence and specific character of *B*-facts nomologically supervene on *A*-facts. Now *B*-facts nomologically supervene on *A*-facts if and only if

> (i) *A*-facts nomologically necessitate *B*-facts,
> (ii) *B*-facts are either downwards identical to *A*-facts, or not downwards identical to *A*-facts, yet
> (iii) nomologically necessarily, there can be no change in any of *X*'s *B*-properties without a corresponding change in *X*'s *A*-properties, and
> (iv) nomologically necessarily, any two beings that are *A*-property indiscriminable are also *B*-property indiscriminable (but not necessarily conversely—in case *B*-facts are not downwards identical to *A*-facts).

As before, the domain of *A*-facts is the supervenience base and the domain of *B*-facts is the supervening domain. And here are the main implications of a nomological supervenience thesis: Fix all the *A*-facts and then you have thereby fixed, with natural or physical necessity (that is, holding in every logical possible world with same kind of physical matter and the same set of natural laws as the actual world), all the *B*-facts. Or otherwise put: Know empirically everything there is to know about the *A*-facts, and you thereby know, empirically, everything there is to know about the *B*-facts.

I want now again to reformulate and update Moore's classical Open Question Argument, although this time as a "synthetic conceivability entails real possibility" non-supervenience argument against *non*-reductive ethical naturalism.

But first I will define non-reductive ethical naturalism. Non-reductive ethical naturalism is the disjunction of non-reductive consequentialism (whether act or rule consequentialism), non-reductive hedonism, non-reductive ethical egoism, and something I will (again) rather inelegantly call "non-reductive ethical hybridism"—

> *Non-Reductive Consequentialism*: All the moral facts are naturally determined by good-consequence facts, but are not nothing but good consequence facts—instead, they are, in some sense that is perhaps a merely conceptual or epistemic sense and not a metaphysical sense, something over and above good consequence facts.
> *Non-Reductive Hedonism*: All the moral facts are naturally determined by positive-pleasure facts, but are not nothing but positive-pleasure facts—instead, they are, in some sense that is perhaps a merely conceptual or epistemic sense and not a metaphysical sense, something over and above positive-pleasure facts.
> *Non-Reductive Ethical Egoism*: All the moral facts are naturally determined by in-my-best-self-interest facts, but are not nothing but in-my-best-self-interest fact—instead, they are, in some sense that is perhaps a merely conceptual or epistemic sense and not a metaphysical sense, something over and above in-my-best-self-interest facts.
> *Non-Reductive Ethical Hybridism*: All the moral facts are naturally determined by some or another hybrid mixture of good-consequence facts, positive-pleasure facts, or in-my-best-self-interest facts, but are not nothing but some or another hybrid mixture of good-consequence facts, positive-pleasure facts, or in-my-best-self-interest fact—instead, they are, in some sense that is perhaps a merely conceptual or epistemic sense and not a metaphysical sense, something over and above some or another hybrid mixture of good-consequence-facts, positive-pleasure-facts, or in-my-best-self-interest facts.

Assuming this disjunctive definition, here is the basic form of my second Existential Kantian Ethics-grounded, reformulated, and updated classical Moorean Open Question Argument, this time against non-reductive ethical naturalism:

> (1) Purportedly, the domain of *B*-facts (= the right) nomologically supervenes on the domain of *A*-facts (= some or another set of natural facts).
> (2) But synthetically conceivably, hence both logically possibly and also really possibly, all the same *A*-facts (= the same set of natural facts) can obtain in another minimally physically duplicated nomologically possible world, but not all the same *B*-facts obtain in that world (= not right in that world)?
> (3) If yes, then *B*-facts (= the right) do not nomologically supervene on *A*-facts (= some or another set of natural facts), hence the thesis that moral facts (= the right) nomologically supervene on some or another set of natural facts commits the second naturalistic fallacy and is false.

Correspondingly, here is my second Existential Kantian Ethics-grounded, reformulated, and updated Open Question Argument, itself, this time as a "synthetic conceivability entails real possibility" non-supervenience argument against non-reductive ethical naturalism:

> (1) Purportedly, the domain of *B*-facts (= the right) nomologically supervenes on the domain of *A*-facts (= either good-consequence facts or positive-pleasure facts or in-my-

best-self-interest facts, or some hybrid mixture of them (= some or another set of natural facts).
(2) But synthetically conceivably, hence both logically possibly and also really possibly, all the same good-consequence facts or positive-pleasure facts or in-my-best-self-interest facts, or some hybrid mixture of them (= the same set of natural facts) can obtain in another minimally physically duplicated nomologically possible world, but not all the same *B*-facts obtain in that world (= not right in that world)?
(3) If yes, then *B*-facts (= the right) do not nomologically supervene on *A*-facts (good-consequence facts or positive-pleasure facts or in-my-best-self-interest facts, or some hybrid mixture of them (= some or another set of natural facts), hence the thesis that moral facts (= the right) nomologically supervene on some or another set of natural facts (= good-consequence facts or positive-pleasure facts or in-my-best-self-interest facts, or some hybrid mixture of them) commits The Naturalistic Fallacy* and is false.

For example: Pick any basic act *A* in the actual world and fix any set of good-consequence facts or positive-pleasure facts or in-my-best-self-interest facts, or any hybrid mixture of them, relative to *A*, and also assume that *A* is prima facie right in the actual world just by virtue of fixing those facts, and also assume that the specific character of *A* is whatever it adventitiously happens to be. It is still synthetically conceivable and therefore both logically possible and also really possible that *A* is morally wrong in some *nomologically possible minimal* good-consequence facts or positive-pleasure facts or in-my-best-self-interest facts or hybrid-mixture-of-them facts *physical duplicate* of the actual world in which the specific character of *A* spontaneously varies. To show this, as before, let us say that *A* is now specifically an attempt arbitrarily to torture some completely innocent person *P* to death, like the Nazis did, for no good reason at all. Even despite being prima facie right in the actual world on consequentialist, hedonistic, egoistic, or hybrid grounds, *A* is always wrong, no matter what. So neither good-consequence facts nor positive-pleasure facts nor in-my-best-self-interest facts, nor any hybrid mixture of them, constitutes the determining ground of rightness, and non-reductive ethical naturalism is false. Again, the determining ground of moral rightness is just the essentially embodied basic intentional act of a rational animal, for better or worse, insofar as it is immanently structured by the Categorical Imperative.

1.5.3 Three Possible Objections, and Six Replies

Here are three possible objections to my ethical anti-naturalist arguments:

(i) analytic conceivability does not entail logical possibility, and synthetic conceivability does not entail real possibility,
(ii) "intuitions are epistemologically useless,"[69] and

(iii) my analysis of basic acts is false.

As to the first objection, I will start with a *negative* reply. It is clear that the only arguments in the recent and contemporary philosophical literature that would suffice to show that analytic conceivability does not entail logical possibility, and also that synthetic conceivability does not entail real possibility, are

(i) that there is no intelligible or defensible analytic-synthetic distinction, as per Quine,
(ii) psychologizing conceivability by reducing it to imaginability, and
(iii) appealing to necessary a posteriori identities and a priori "illusions of contingency."

Correspondingly, then, first, elsewhere I have thoroughly criticized Quine's critique of the analytic-synthetic distinction, thereby opening up a new place in theoretical space for an intelligible and defensible positive theory of the distinction.[70] Second, elsewhere I have also argued against logical psychologism, both by refining Husserl's classical arguments and also by developing a new non-Husserlian argument based on the notion of logical supervenience.[71] Third, elsewhere I have also argued against the very idea of the necessary a posteriori, showing it to be based on fundamental confusions about the a priori/a posteriori distinction and about how many distinct types of propositions can be expressed in a given speech-context, and correspondingly proposing its elimination.[72] Compatibly with that, David Barnett has worked out an independently-motivated argument against the very idea of necessary a posteriori identities in meta-ethics.[73]

Here, now, is a *positive* reply to the first objection. I think that it can be shown by the following line of argument that analytic conceivability does indeed entail logical possibility, and also that synthetic conceivability does indeed entail real possibility. First, I provide an intelligible and defensible positive theory of the analytic-synthetic distinction.[74] Second, I adopt a cognitivist modal framework according to which a logically (or analytically) possible world is a maximal consistent set of different conceivable ways the actual world could have been, and a really (or synthetically) possible world is a maximal consistent set of conceivable ways the actual world could have been, as constrained by the essentially-non-conceptually-represented underlying spatiotemporal, causal-dynamic, and mathematical structures of the actual world.[75] And third, I show that analytic conceivability (via a priori conceptual competence) cognitively accesses logical possibility, and also that synthetic conceivability (via a priori conceptual competence and a priori essentially non-conceptual content) cognitively accesses real possibility.[76]

As to the second objection, I have elsewhere argued not only in a negative way against logical psychologism, but also in a positive way for the self-evidence, necessary truth, and essentially reliable justification of at least *some* rational intuitions in logic.[77] And, correspondingly, I have elsewhere argued not only in a negative way against mathematical psychologism, and against skepticism and eliminativism about rational intuitions too, but

also in a positive way for the self-evidence, necessary truth, and essentially reliable justification of at least some rational intuitions in mathematics, logic, and philosophy more generally. [78] In short, *I have done my level best to defend intuitions*.[79]

Finally, as to the third objection, elsewhere I have argued for the truth of my analysis of basic intentional acts in the context of a fundamental theory of the mind-body relation and mental causation.[80]

So I think that I am rationally entitled to reject all three of the objections to my arguments. Or at the very least, I can reasonably claim a philosophical draw: I have done *my* burden-of-proof-shouldering homework; so, on their side, my critics now need to re-group and re-think.

In a nutshell, the core critical point lying behind my two Existential Kantian Ethics-grounded, reformulated, and updated versions of Moore classical Open Question Argument against ethical naturalism is this: We can both analytically conceivably and also synthetically conceivably (and thus both logically possibly and also really possibly) vary the specific prima facie moral character of basic intentional acts in the actual world over these logically or nomologically possible worlds:

> (i) minimal good-consequence fact physical duplicate worlds (contra naturalistic act or rule consequentialism),
> (ii) minimal positive-pleasure fact physical duplicate worlds (contra naturalistic hedonism),
> (iii) minimal in-my-best-self-interest fact physical duplicate worlds (contra naturalistic ethical egoism), and also
> (iv) minimal hybrid good-consequence fact, positive-pleasure fact, or in-my-best-self-interest fact physical duplicate worlds (contra naturalistic ethical hybridism).

From this, we immediately derive the non-supervenience of moral facts on any naturalized ethical facts, whether logical supervenience or nomological supervenience, and thereby also immediately derive the falsity of either reductive or non-reductive ethical naturalism. So, by virtue of my two reformulated and updated versions of Moore's Open Question Argument, it follows that reductive and non-reductive ethical naturalism alike are false.

Furthermore, it should be noted that the specific prima facie moral characters of basic intentional acts in the actual world can be both analytically conceivably and also synthetically conceivably (and thus both logically possibly and also really possibly) spontaneously varied over any set of natural facts whatsoever. This is precisely because the specific prima facie moral characters of basic intentional acts vary according to free willing, which is radically underdetermined by natural facts of *any* kind. Those are Goethe's and Wittgenstein's deep points in the *Faust* and *Tractatus* texts I quoted as the epigraphs of this section.

Free willing, according to the account I presented and defended in *Deep Freedom and Real Persons*, is a pre-reflectively conscious and also self-conscious hierarchical desire-

structuring, whereby an essentially embodied human person can always spontaneously choose to adopt an attitude or change her attitude at the level of either pre-reflectively conscious or self-conscious first-order effective desires and trying. And this in turn necessarily determines the specific moral character of her basic intentional act, independently of natural facts of any kind, including deterministic natural facts and stochastic law-governed indeterministic natural facts. But this specific moral act-character has irreducible moral properties that are also thereby determined independently of natural facts of any kind, again including deterministic natural facts and stochastic law-governed indeterministic natural facts, by virtue of the immanent structuring of all basic intentional acts of real human persons by the Categorical Imperative.

At bottom, then, ethical naturalism is false just because, although essentially embodied human persons are fully natural beings—as real human persons, they are rational human minded *animals*—nevertheless they are *not* strictly determined by contingent natural facts. So, in the beginning was The Basic Act. As a direct consequence, "the world of the happy" is a fully *natural* but also *non-naturalizable* world.

1.6 WHY, DEEP IN YOUR HEART, YOU ARE AN EXISTENTIAL KANTIAN ETHICIST

Lastly, and perhaps above all, in this book I hope to be able to convince you that you are already, have always been, and always will be—at least implicitly and pre-reflectively, even if not also explicitly and self-consciously—an existential Kantian ethicist, in the following two senses.

First, it is rationally intuitive or self-evident to you that in each of the two pairs of cases presented in section 1.3 above, option (ii) is the morally better choice. Here, again, are those options:

FIRST PAIR

Option (i): a guaranteed life of shallow happiness for herself, even at the cost of her treating someone else either as a mere means to other's ends or as a mere thing, that is, like the Nazis treated people, provided that no one else ever finds out (ethical egoism).
Option (ii): a guaranteed life of pointless suffering for herself, just because she refuses to treat someone else as a mere means or a mere thing, like the Nazis treated people, even if it were guaranteed that no one else would ever find out that she had treated someone that way (anti-egoism).

SECOND PAIR

Option (i): guaranteed lives of shallow happiness for almost everyone else, at the cost of her treating someone else as a mere means or mere thing, like the Nazis treated people, provided that no one else ever finds out (act consequentialism).
Option (ii): exactly the same sort of actual lives for everyone else (that is, no shallowly happier lives for anyone else) plus a guaranteed life of pointless suffering for herself, just because she refuses to treat someone else as a mere means or mere thing, like the Nazis treated people, even if it were guaranteed that no one else would ever find out that she had treated someone that way (anti-act consequentialism).

Initially, it may not seem intuitively self-evident to you that it is intuitively self-evident to you that option (ii) is always the morally better choice. But here is the crucial caveat. The question to be answered by your activated capacity for rational intuitive judgment is not whether you think that you yourself, in your current "human, all too human" condition, will actually choose option (ii). Instead, the question to be answered by your activated capacity for rational intuitive judgment is whether you think

(i) that it is at least really possible for you to have chosen it, and also
(ii) that had you chosen it, then your choice would have been the morally better choice.

In other words, your activated capacity for rational intuitive judgment is being asked to evaluate not what is actually psychologically possible for you right now, but instead what is really possible for you in a counterfactual situation, even if you actually do not choose option (ii). Once you make that distinction, it will be intuitively self-evident to you that option (ii) is the morally better choice.

Second, as long as you are still alive-and-kicking,[81] it is never too late for you to achieve principled authenticity and a fully meaningful life, at least partially or to some degree. With good luck, human happiness could also flow from this full meaningfulness, and then you would be happy in all basic senses of that term—including deep and shallow happiness alike—to that extent too. And to the extent that this is rationally humanly possible, then that would be the best of all really possible lives for creatures like us.

But even with quite a lot of bad luck and pointless suffering, as long as you had achieved principled authenticity at least in part or to some degree, then you would still have had a logically and morally excellent and cogent life to that extent, just like a logically consistent, valid, and sound argument that necessarily guarantees the truth of its conclusion. Camus says that we can imagine Sisyphus to be "happy." But I think that it would be more accurate and illuminating to say that we can imagine Sisyphus to be authentic.

Even more precisely, however, and now to coin a phrase, I hold these three truths to be self-evident—

> (i) that every rational human animal or real human person belongs to The Realm of Ends,
> (ii) that each is innately endowed by our equally rational and animal nature with certain unalienable desires, capacities, and needs, and
> (iii) that among these are: the desire for life; the innately specified online capacities for Kantian autonomy and for Kierkegaardian purity of heart or wholeheartedness; and the need for, and the pursuit of, deep and not merely shallow happiness.

In turn, the high-bar, maximal, or ideal rational normative standard of principled authenticity is what ultimately controls all of these. So according to Existential Kantian Ethics, every rational human animal or real human person who has reached the stage of Kantian or higher-level personhood, at least implicitly believes, deep in her heart, the following four things—

> (i) that there is something that is truly worth living for and also truly worth dying for, which is the highest or supreme moral value, and thereby also the ultimate high-bar, maximal, or ideal standard of rational normativity,
> (ii) that this highest or supreme moral value and standard is shared by all rational human animals or real human persons in common, even if they are not self-consciously or self-reflectively aware of it (for example, all or most normal, healthy children are self-consciously or self-reflectively unaware of it),
> (iii) that this highest or supreme moral value and standard, as pursued by that real human person in this thoroughly nonideal natural and social world, is also as profoundly individual as the unique real personal life of each rational human animal, and finally
> (iv) that the realization of principled authenticity, at least partially or to some degree, is this highest or supreme moral value and standard.

Of course we never, not even in principle, manage to live up *fully* to this highest or supreme moral value and standard. That would be an impossible perfection for "human, all too human" animals like us. Although we are clearly somewhat perfectible, we are at the same time so manifestly imperfect, and so frequently morally awful, both to others and to ourselves, and also the natural world and social all around us is so thoroughly nonideal, as to make the possibility of our fully achieving moral perfection an obvious non-starter. But then we might naturally wonder: How can it be intelligible that our moral goals so radically outreach our moral abilities? In effect, it is The Riddle of the Sphinx revisited:

> Which creature in the morning goes on four legs, at mid-day on two, and in the evening upon three, and the more legs it has, the weaker it is?

Oedipus's answer to the original riddle, namely, that it is *the human being* who crawls on four legs as a baby, walks on two legs as an adult, then hobbles three-leggedly with a cane as an oldster, and is weakest precisely insofar as it should be strongest, is essentially my

answer too. We are nothing more and nothing less than the animals with the very highest moral goals and the very lowest moral success rates—that is, the animals capable of evil and suffering. And it is precisely because our moral goals are so very high, that our moral success rates are so very low, and correspondingly that our perverse ability for both banal evil and near-satanic evil alike[82] is so very effective, and that our suffering is so very intense and widespread.

Perhaps the most achingly beautiful moment in perhaps the most achingly beautiful film ever made, Yasujiro Ozu's *Tokyo Story*, occurs when one of the characters says to another: "Isn't life disappointing?," and the other replies, with an ineffably sad smile on her face, "Yes it is." In other words, our having this highest or supreme moral value innately specified within us is what partially defines us, but only as inseparably taken together with our "human, all too human" limitations. How it is that we might be able to come to terms with our own inherent finitude and with our own sheer, bent humanness, in the face of our deepest commitment to the high-bar, maximal, or ideal rational normative standard of principled authenticity, is the last and supremely hardest-to-grasp element in Existential Kantian Ethics. But by way of a preview, I can say this. Taking another cue from Ozu, whose

> tombstone bears the single character for *mu*—an aesthetic word, a philosophical term, one which is usually translated as "nothingness" but which suggests the nothing that in Zen philosophy, is everything[,][83]

I will claim that it all has essentially to do with *the morality of one's own death*. I will return to this in chapter 6.

Chapter 2

LIVING WITH CONTRADICTIONS: NONIDEAL KANTIAN ETHICAL THEORY

A conflict of duties would be a relation between them in which one of them would cancel the other (wholly or in part)…. But since duty and obligation are concepts that express the objective practical necessity of certain actions and two rules opposed to each other cannot be necessary at the same time, if it is a duty to act in accordance with one rule, to act in accordance with the opposite rule is not a duty but even contrary to duty; so a collision of duties and obligations is inconceivable. However, a subject may have, in a rule he prescribes to himself, two grounds of obligation, one or the other of which is not sufficient to put him under obligation, so that one of them is not a duty. (*MM* 6: 224)

A foolish consistency is the hobgoblin of little minds, adored by little statesmen and philosophers and divines.[84]

That an act qua fulfilling a promise, or qua effecting a just distribution of good, or qua returning services rendered, or qua promoting the virtue or insight of the agent, is prima facie right, is self-evident; not in the sense that it is evident from the beginning of our lives, or as soon as we attend to the proposition for the first time, but in the sense that when we heave reached sufficient mental maturity and have given sufficient reflection to the proposition it is evident without any need of proof, or evidence beyond itself. It is self-evident just as a mathematical axiom, or the validity of a form of inference, is evident. The moral order expressed in these propositions is just as much a part of the fundamental nature of the universe (and, we may add, of any possible universe in which there are moral agents at all) as the spatial or numerical structure expressed in the axioms of geometry or arithmetic. In our confidence that these propositions are true there is involved the same trust in our reason that is involved in our confidence in mathematics; and we should have no justification for trusting it in the latter sphere and distrusting it in the former.[85]

Hide what you have to hide
And tell what you have to tell
You'll see your problems multiplied
If you continually decide
To faithfully pursue
The policy of truth[86]

2.1 How Nonideal Can A World Be?

Strictly and narrowly speaking, in recent and contemporary Anglo-American professional academic philosophy, "nonideal theory" is *political* theorizing under the assumption that compliance to *principles of justice* is inherently not strict. But there is a broader and deeper sense of "nonideal theory" that is *ethical* theorizing under the assumption that compliance to *moral principles* is inherently not strict. It is this broader and deeper sense that I am particularly interested in. More precisely, in this chapter I want to work out the basics of *nonideal Kantian ethical theory*.

Sadly, there are nonideal worlds, and then there are nonideal worlds. How nonideal can they be? A moral contradiction, or moral dilemma, may be defined as a situation *S* in which a person *P*, given her moral principles, ought to choose or do some act *A* in *S* and also ought to choose or do some act *B* in *S*, such that *P*'s choosing or doing *A* entails her choosing or doing not-*B* in *S* and also *P*'s choosing or doing *B* entails her choosing or doing not-*A* in S. Sartre's famous case of the boy and his mother is of course one example:

> The boy was faced with the choice of leaving for England and joining the Free French Forces—that is, leaving his mother behind—or remaining with his mother and helping her carry on…. Who could help him choose? … Who can decide a priori? Nobody. No book of ethics can tell him. The Kantian ethics says, "Never treat any person as a means but as an end." Very well, if I stay with my mother, I'll treat her as an end and not as a means; but by virtue of this very fact, I'm running the risk of treating the people around me who are fighting, as means; and conversely, if I go to join those who are fighting, I'll be treating them as an end, and, by doing that, I run the risk of treating my mother as a means. If values are vague, and if they are always too broad for the concrete and specific case that we are considering, the only thing left is to trust our instincts. That's what this young man tried to do; and when I saw him, he said, "In the end, feeling is what counts. I ought to choose whatever pushes me in one direction…" But how is the value of that feeling determined? What gives his feeling for his mother value? Precisely the fact that he remained with her.[87]

But here is an even more poignant, real-world, and, for me, close-to-home example, involving a Kantian philosopher I knew and studied philosophy with. Other things being equal, you ought to preserve rational human life; and other things being equal, you also ought to prevent the suffering of other real persons if you can, especially the suffering of those you love most; and then your suffering sibling, parent, or life-partner lucidly asks you to assist in his or her suicide:

> The story was brief, tragic and haunting. A brilliant philosophy professor, Stephan Körner, had been found dead with his wife Edith, an NHS pioneer who had just been diagnosed as having terminal cancer. Instead of being divided by disease the couple chose

> to be united in death, taking a lethal overdose and breathing their last in each other's arms at their Bristol home.[88]

Moral contradictions or moral dilemmas in the sense I have just spelled out can originally derive from one or more moral principles. Sartre's case is a two-principle case, as is the Stephan and Edith Körner case. In *Sophie's Choice*, by contrast, Sophie is originally committed to a single moral principle which says that she ought to protect both of her children; yet the deeply unlucky context of action also brings it about that under her original moral principle Sophie is also committed to a sub-principle of selecting one of her children to be killed by the Nazis, since otherwise both children will be killed by them. In any case, I will call any nonideal world in which moral contradictions or dilemmas—whether derived from one or many moral principles—do in fact occur, and also occur with alarming frequency, a *thoroughly* nonideal world. Our actual natural and social world, it seems clear, is a thoroughly nonideal world. Thoroughly nonideal worlds, then, are those intentional-act-worlds in which not only is it the case that compliance to moral principles is inherently not strict, but also moral dilemmas all-too-frequently happen.

Consider now a real "human, all-too-human" person of good will, living in our thoroughly nonideal natural and social world. When, as is almost inevitable in this world, her principles come into real conflict with one another in some morally unlucky situation, then this real human person of good will must wholeheartedly choose the lesser of several evils in that context of action, and also take complete responsibility, with no excuses, for something over which she had no control whatsoever—namely, the brute contingent fact of conflicting principles in that act-context. That brute contingent fact is also the morally tragic fact that every one of her choices in that situation will involve a violation of at least one of her principles. There is no way out. She must wholeheartedly choose, and then bravely and stoically take an awful hit.

This is what I call *The Kant-Sartre Insight*.[89] In turn, I want to use The Kant-Sartre Insight as a rational-intuitive guide, or philosophical pole-star, to working out the basics of nonideal Kantian ethical theory. In so doing, I will develop a new interpretation and conservative extension of the highly influential and equally notorious ethical theory laid out by Kant in the *Groundwork*, the *Critique of Practical Reason*, the *Metaphysics of Morals*, and *Religion within the Boundaries of Mere Reason*. Borrowing a famous phrase from Emerson, I call this new interpretation and conservative extension *The No-Foolish-Consistency Interpretation*. If I am correct, The No-Foolish Consistency Interpretation provides a unified solution to three classical problems in Kantian ethics:

> (i) *the problem of universalizability*, or the apparent epistemic indeterminacy of tests for the generalizability and consistency of moral principles,[90]
> (ii) *the problem of rigorism*, or the apparent over-strictness, apparent overgeneralization, and apparent overly-extended strictly universal scope, of moral principles,[91] and above all,

(iii) *the problem of moral dilemmas*, or the apparent inconsistency between equally legitimate absolutely universal moral principles.[92]

In solving the third of these problems, we will find that Kant himself was clearly and even scandalously mistaken about the semantic structure and normative implications of his own theory of moral principles, and also that W.D. Ross was much closer to the truth about these matters, although still not quite adequate to the phenomena. Ross stood on the shoulders of Kant, and saw a little further than Kant did. My hope is that by standing, *Cirque du Soleil*-wise, on the shoulders of these two giants of Kantian ethics, I will be able to see just a little further than either of them did.

2.2 THE SKINNY LOGIC ND THE FAT SEMANTICS OF MORAL PRINCIPLES IN EXISTENTIAL KANTIAN ETHICS

The background conceptions of logic and semantics that I am using in this chapter and throughout this book as a meta-ethical foundation, include

(i) Kant's "pure general logic" (*CPR* A50-57/B74-79) (*JL* 11-20, 91-150), or what we would now think of as second-order intensional monadic logic, that is, classical truth-functional logic together with a restricted predicate logic employing quantification into and also over one place predicates only, and quantifying over individuals (= first-order monadic logic[93]), but also quantifying over the Kantian concepts or finegrained intensions expressed by one-place predicates (= second-order intensional monadic logic),[94] and
(ii) Kant's "transcendental logic" (*CPR* A55-57/B79-82), or what we would now think of as a finegrained intensional possible worlds semantics of propositions together with what I call *Kantian modal dualism*.

Now according to the robust semantics of Kantian modal dualism, in turn, there are two irreducibly and essentially different kinds of necessary truth:

(i) *analytic necessity*, which is a priori necessary truth in virtue of conceptual content, always taken together with some things in the world beyond conceptual content, although never in virtue of those worldly things, that is, the necessity that flows from concepts, and (2) *synthetic necessity*, which is a priori necessary truth in virtue of things in the world beyond conceptual content, that is, truth in virtue of pure intuitional and imaginational content representing the underlying non-empirical intrinsic spatiotemporal, causal-dynamic, and mathematical immanent structures of matter in the actual world, always taken together with some conceptual content, although never in virtue of conceptual content, that is, the necessity that flows from things in the world.[95]

So, in other words, the background conceptions of logic and semantics I am using as a dual logico-semantic meta-ethical foundation for Existential Kantian Ethics jointly provide, at one and the same time, for a somewhat *thinner and more minimalist* conception of pure logic than standard classical logic in the Frege-Russell-Carnap-Tarski-Quine tradition (aka "elementary logic") and also for a somewhat *thicker and more robust* conception of semantics than standard classical semantics in the mainstream Frege-Russell-Carnap-Tarski-Kripke tradition. I argued in detail and at length in *Kant the Foundations of Analytic Philosophy* and again in *Cognition, Content, and the A Priori*, that although Kant is almost universally criticized for having a skinnier logic and a fatter semantics than most logicians and semanticists in the mainstream Frege-Russell-Carnap-Tarski-Quine-Kripke tradition are prepared to accept, nevertheless, there are very good reasons to think that they are wrong, and Kant was right.[96] As controversial as those claims are, in order not to overburden the present book, I will not attempt to re-argue those claims here and will simply assume the soundness of my earlier arguments.

Still, I do want to emphasize right from the outset that Existential Kantian Ethics does indeed presuppose a special non-classical logic and also a special non-classical semantics, and also that *if* we take this skinnier logic and that fatter semantics explicitly into account, *then* our conception of Kantian ethical theory will be significantly deepened and strengthened, as per The No-Foolish-Consistency Interpretation.

According to Kant in the *Groundwork* and also later in the *Critique of Practical Reason* (*CPrR* 5: 19-28), the *Metaphysics of Morals* (*MM* 6: 211-227), and *Religion within the Boundaries of Mere Reason (Rel* 6: 3-5, 20-50), and also according to Existential Kantian Ethics, morality is grounded on a set of strictly and unconditionally universal a priori normative meta-principles which are categorically binding on all rational beings, and more specifically are categorically binding on all rational human animals, insofar as all rational human animals have

> (i) a "will" (*Wille*), which is an innate psychological capacity for rational desiring, or practical justification in terms of either non-instrumental reasons (namely, the Categorical Imperative in its four or five analytically equivalent formulations) or instrumental reasons, and also
> (ii) a "power of choice" (*Willkür*), which is an innate psychological capacity for effective desiring, or causally efficacious conscious motivation to choice and action.

Otherwise put, the *Wille* is a *legislative* practical capacity that generates, recognizes, and is more generally reasons-sensitive to principles and imperatives, whereas the *Willkür* is an *executive* practical capacity that enacts and implements principles and imperatives by means of reasons-sensitive conscious conations, drives, or impulses. Together, the faculties of *Wille* and *Willkür* jointly constitute a dual faculty for rational desire-based choice. In turn, the *Wille* has

> (i) a higher proper part (*pure practical reason*, that is, a power for non-instrumental reasoning) that generates, recognizes, and reasons with categorical imperatives, and also
> (ii) a lower proper part (what I will dub *impure practical reason*, that is, a power for instrumental reasoning) that generates, recognizes, and reasons with hypothetical imperatives.

So, to summarize, according to Existential Kantian Ethics, the overall structure of the human will or faculty of desire looks like this:

> Human Will or Faculty of Desire (*Begehrungsvermögen)*:
> higher part = faculty of practical reason or will proper (*Wille)*:
> higher part = pure or non-instrumental reason
> lower part = impure or instrumental reason
> lower part = power of choice (*Willkür)*

Moreover, according to the nonideal Kantian theory of moral principles that I am developing here, we need to distinguish very sharply between

> (i) *absolutely universal and objective moral meta-principles*, which are strictly and unconditionally universal and objective a priori normative rules binding on all rational beings, including all rational human animals or real human persons,
> (ii) *first-order substantive ceteris paribus objective moral principles* (aka "fairly universal and objective moral principles"—see section 1.2 above), which tell us what we ought to do, other things being equal, and are binding on all rational human animals or real human persons in any set of circumstances, provided that certain favorable background conditions obtain, and finally
> (iii) *moral duties*, which are first-order objective moral principles that are also agent-centered obligations.

In given act-contexts, moral agents can find that other things really *are* equal. So moral duties are first-order substantive ceteris paribus objective moral principles with agent-centered application, under absolutely universal and objective moral meta-principles.

A first-order substantive ceteris paribus objective moral principle is essentially the same as what Kant calls a "ground of obligation" (*MM* 6: 224). A ground of obligation is a morally sufficient reason for choosing-and-acting or for refraining, other things being equal. Similarly, to use an everyday analogy, your mother, father, or kindergarten teacher can tell you what you ought to do or not do, other things being equal, and s/he might be completely right. But because our actual natural and social world is a thoroughly nonideal world, *other things really might not be equal in any given actual act-context*; and, correspondingly, because the ceteris paribus condition therefore *really might not be satisfied in that actual act-context*, it does not automatically follow that you are obligated to do what your mother, father, or teacher rightly tells you that you ought to do—unless, in

that act-context, things *really are* equal, and you yourself *really can* do it. So in order to be a moral duty, a first-order substantive ceteris paribus objective moral principle has to have adequate agent-centered force in an actual act-context, and this depends in part on the way the world and other people just contingently really happen to be, quite independently of the agent herself, as well as depending in part on the actual agent herself and her agential capacities in that actual act-context.

This, in turn, shows us how to interpret the well-known Kantian principle that "ought implies can" (*CPR* A548/B576, A807/B835) (*MM* 6:380). As Robert Stern has pointed out, the Kantian version of "ought implies can" does *not* mean that "nothing can be right that we are incapable of achieving," but *instead* means that "we cannot be obliged to do what is right unless we are capable of acting in that way."[97] As we all know, there is significant contingent variability in how our basic shared agential capacities are actually realized in different real human persons, in different act-contexts. Thus moral duties obligate us to do what some moral principles tell us we ought to do, other things being equal—that is, leaving out contingent conditions in act-contexts. But if we reintroduce contingent conditions in act-contexts, then we might be morally obligated, although it is not necessarily the case that we will be morally obligated; for we do not have a duty in each actual act-context, but rather only in some actual act-contexts. I will come back to this point in section 2.2 below.

Necessarily, every moral duty is also a first-order substantive ceteris paribus objective moral principle, but not every first-order substantive ceteris paribus objective moral principle is also a moral duty. This is because there can be real conflicts between first-order substantive ceteris paribus objective moral principles, even in cases in which an agent has one and only one moral duty:

> A subject may have, in a rule he prescribes to himself, two grounds of obligation …, one or the other of which is not sufficient to put him under obligation, so that one of them is not a duty. (*MM* 6:224)

Indeed, while there can be real conflicts of first-order substantive ceteris paribus objective moral principles, there cannot be conflicts of moral duties, as a matter of analytic a priori necessity: "a collision of duties and obligations is inconceivable" (*MM* 6: 224). Thus the distinction between first-order substantive ceteris paribus objective moral principles and moral duties captures the essence of what Ross was driving at in his famous distinction in *The Right and the Good* between "prima facie duties" and "actual duties," but without the strange consequence that a given moral principle can *sort-of* be my moral duty without its also *really* being my moral duty.

In *The Right and the Good*, Ross argues that we have rational, self-evident, non-inferential, infallible a priori intuitions about an irreducibly plural class of co-basic moral principles, the *prima facie* duties, which include the seven duties of "fidelity," "reparation,"

"gratitude," "justice," "beneficence," "self-improvement," and "non-maleficence."[98] These seven principles, purportedly, are knowable by any mature, reflective rational human animal. *Prima facie* duties are sharply distinguished from actual duties, or duties proper, that

> (i) are the objectively real moral obligations binding on moral agents or persons in particular act-contexts, and
> (ii) are objectively determined by their being the moral principles that, in that act-context, have the greatest balance of prima facie rightness over prima facie wrongness, of all possible acts for that agent in that context, when the act is taken in "its whole nature."[99]

At the same time, however, according to Ross, it is not possible rationally to intuit, or authentically to know, actual duties—at best, it is possible to cognize actual duties with "right opinion," and not sufficiently justified true belief,[100] that is, *essentially reliable justified true belief*, or what elsewhere I call *High-Bar justified true belief*, whereby there is an intrinsic connection between the warranting evidence for belief, as delivered by our properly functioning cognitive capacities or mechanisms, and the truth.[101]

In any case, Ross has three main goals in combining the classical theory of rational intuition with his theory of *prima facie* duties vs. actual duties.

First, he wants to provide a secure, realistic, and a priori but also non-monistic foundation for moral theory.

Second, he wants to accommodate the obvious empirical fact of conflicts of duties—moral contradictions or moral dilemmas—that seem to arise directly from the foundational fact of a plurality of basic duties together with our actual "human, all too human" existence in this thoroughly nonideal natural and social world.

And third, he wants to incorporate some measure of commonsensical or real-world fallibilism about our moral judgments in particular contexts in this world.

Ross's moral intuitionism is thereby designed precisely in order to accommodate the thoroughly nonideal character of our moral lives. But his intuitionism is also philosophically notorious. Correspondingly, here are the three classical critical objections to Ross.

First, Ross's postulation of a mysterious and "queer" (in Mackie's sense[102]) faculty for intuitively knowing the prima facie duties has no independent plausibility or empirical support whatsoever.

Second, Ross's infallibilism about moral intuition seems to fly in the face of the highly plausible thesis of fallibilism about a priori knowledge, as well as fallibilism about empirical knowledge.

And third, the obvious empirical fact of widespread disagreement, even amongst mature, reflective rational human animals, about precisely which moral principles are true

and which are false seems to undermine completely Ross's claim that even some moral principles are known intuitively with self-evidence.

In addition to these three classical worries, I also have a non-classical, fourth critical objection that in certain respects is similar to John Rawls's main worry about Ross's theory, to the effect that Ross cannot ultimately avoid a theory of the lexical ordering and weighting of the supposedly equally morally binding, lexically unordered, and unweighted prima facie duties.[103] More precisely, my non-classical, Rawls-inspired critical objection can be posed as a dilemma:

> Ross's theory of prima facie duties, on the one hand, explicitly postulates an irreducible *pluralism* of basic moral principles; yet on the other hand, he implicitly presupposes *a monistic deontological scale* in explicating the advance from prima facie duties to actual duties—otherwise how could there be an objective determination of one moral principle's being the one which, in a given context, expresses the greatest balance of prima facie rightness over prima facie wrongness, of all possible acts for that agent in that context, when the act is taken in "its whole nature"?

In *Cognition, Content, and the A Priori*, I respond to the three classical worries about Ross's moral intuitionism, at least by implication, by developing a contemporary Kantian theory of rational intuition that is an extension of a contemporary Kantian theory of mathematical intuition, and also includes moral intuition as a sub-case.[104] In the present context, however, I want to respond directly only to the non-classical, fourth objection to Ross's moral intuitionism by extending the notion of *structuralism* from mathematics to morality.

Mathematical structuralism, as an explanatory metaphysical thesis in the philosophy of mathematics—defended, for example, by Stewart Shapiro,[105] and in another way by Charles Parsons[106]—says that mathematical entities (for example, numbers or sets) are not ontologically autonomous or substantially independent objects, but instead are, essentially, positions or roles in a mathematical structure, where a mathematical structure is a complete set of formal relations and operations that defines a mathematical system. What counts as an individual object of the system is thereby uniquely determined by the system as a whole. That is, any such individual object is identical to whatever possesses a specific set of intrinsic structural system-dependent properties. In a text quoted as one of the epigraphs for this chapter, it seems clear enough that Ross himself had a moral structuralist idea in mind:

> The moral order expressed in these propositions is just as much a part of the fundamental nature of the universe (and, we may add, of any possible universe in which there are moral agents at all) as the spatial or numerical structure expressed in the axioms of geometry or arithmetic.

But he never systematically developed or elaborated that important thought.

Nevertheless, Ross's important thought can be unpacked and effectively deployed within the framework of Existential Kantian Ethics. So, standing on Ross's shoulders, here are the six basic ideas behind my existential Kantian version of moral structuralism.

First, there is *a three-levelled hierarchy of moral principles*, not a "flat" or non-hierarchical set of moral principles, as is usually assumed to be the case.

Second, moral principles are not ontologically autonomous, substantially independent, "atomic" semantic or normative objects, but instead are, essentially, positions or roles in a *moral structure*, where a moral structure is a complete set of semantic relations and normative forces that defines a moral system of principles.

Third, the semantic content and normative force of any individual moral principle is thereby determined by the moral system as a whole—that is, *any such individual principle is identical to whatever possesses a specific set of intrinsic structural system-dependent properties*.

Fourth, completely convincing, intrinsically compelling, or self-evident moral intuition *applies only to the top level in the hierarchy, which are procedural meta-principles*, and neither to intermediate-level first-order substantive ceteris paribus moral principles, nor to bottom-level actual duties.

Fifth, the rational advance from the completely convincing, intrinsically compelling, or self-evidently intuited top-level meta-principles to the intermediate-level first-order substantive ceteris paribus principles to the bottom-level actual duties is *a process of cognitive and volitional construction*.

And finally, sixth, real conflicts of first-order substantive ceteris paribus moral principles at the intermediate level of the hierarchy *are automatically resolved by a special set of level-theoretic structural constraints, taken together with one other moral meta-principle called The Lesser Evil Principle*, which collectively fully preserve the absolutely universal objective truth and reality of the authoritatively-intuited meta-principles at the top level of the hierarchy.

As the rest of this chapter rolls out, I will deploy these six moral structuralist ideas against the backdrop of Existential Kantian Ethics, in order to capture The Kant-Sartre Insight and also express the full explanatory power of nonideal Kantian ethical theory.

Moral principles, whether absolutely universal and objective moral meta-principles, first-order substantive ceteris paribus objective moral principles, or moral duties, should also be sharply distinguished from *moral judgments*, which are constructive applications of objective moral principles in particular act-contexts. Indeed, the confusion between objective moral principles and moral judgments is perhaps the most persistent fallacy in both classical and contemporary interpretations of Kant's ethics and Kantian ethics alike. It is one thing to determine the logico-semantic structure and normative implications of a given objective moral principle, and another very different thing to figure out in the thick

of things how a given objective moral principle is to be deployed or instantiated—that is, constructively applied—in a given actual act-context.

Now the thesis of *constructivism*, whether inside or outside ethics, says that human minds and human agents play active, basic roles in determining and generating the content of all beliefs, truths, knowledge (especially including the knowledge of language), desires, volitions, act-intentions, and objective logical or moral principles. Corespondingly, Kantian constructivism in the theory of mental content, cognition, and knowledge (aka *Erkenntnistheorie*) says that innately-specified rules essentially constrain the process by which human minds determine and generate mental representations of a manifest world that must also structurally conform to the formal constitution of their cognitive faculties.[107] And finally, Kantian constructivism *in ethics* says that a fundamental conception of the rational human agent essentially constrains the process by which agents determine and generate first-order substantive objective moral principles.[108]

Given this backdrop, every moral judgment constructively presupposes one or more objective moral principles; but the specific character and general properties of moral judgments cannot be automatically extended to objective moral principles, nor can the specific character and general properties of objective moral principles be automatically extended to moral judgments, since the mind-driven and agent-driven constructive process necessarily intervenes and mediates between the two. I will come back to the important distinction between objective moral principles and moral judgments in section 2.3 below.

In any case, assuming for the purposes of my argument at least the intelligibility of a sharp fourfold distinction between

(i) absolutely universal and objective moral meta-principles,
(ii) first-order substantive ceteris paribus objective moral principles,
(iii) moral duties, and
(iv) moral judgments,

all of which are projected into the larger theoretical frameworks of Kantian constructivism and Existential Kantian Ethics, I want now to address the three classical problems of universalizability, rigorism, and moral dilemmas.

2.3 How to Solve the Universalizability and Rigorism Problems

In the *Groundwork*, Kant provides four (or alternatively, depending on how finegrained one wants the theory of basic moral meta-principles to be, five) distinct formulations of the Categorical Imperative:

The Formula of Universal Law (aka FUL):

> Act only on that maxim by which you can at the same time will that it should become a universal law. (GMM 4: 421)

[Alternative Formulation: The Formula of the Universal Law of Nature:

> Act as though the maxim of your action were to become by your will a universal law of nature. (GMM 4: 421)]

The Formula of Humanity as End-in-Itself (FHE):

> 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. (GMM 4: 429)

The Formula of Autonomy (FA):

> The supreme condition of the will's harmony with universal practical reason is the Idea of the will of every rational being as a will that legislates universal law. (GMM 4: 431)

The Formula of the Realm of Ends (FRE):

> Never .. perform any action except one whose maxim could also be a universal law, and thus .. act only on a maxim through which the will could regard itself at the same time as enacting universal law. (GMM: 433)

Each of the formulas of the Categorical Imperative is a procedural moral meta-principle that tells us *how to select first-order moral principles*. On my interpretation of Kant's theory of moral principles, there is also a *lexical* ordering relation between the Formula of Universal Law/Formula of the Universal Law of Nature and the other three formulas of the Categorical Imperative, considered as a single three-membered set. In other words, the Formula of Universal Law/Formula of the Universal Law of Nature is a *formal presupposition* of the other three procedural moral meta-principles, hence it always logically, semantically, and normatively precedes the three of them, taken as a group. More precisely, the Formula of Universal Law says that nothing will count as an objective moral principle, and in particular nothing will count as a "maxim," unless that objective moral principle or maxim consistently generalizes. Now according to Kant, a maxim is a "principle of volition" (*GMM* 4:400) or act-intention in some or another act-context. So the Formulua of Universal Law, as the formal presupposition of all procedural moral meta-principles, says that nothing will as an objective moral principle, and in particular nothing

will count as a morally permissible objective principle of volition or act-intention in any act-context, unless it consistently generalizes.

The Formula of the Universal Law of Nature, as I am understanding it, is just a specification of the Formula of Universal Law, which in turn says that nothing will count as an objective moral principle, and in particular nothing will count as a morally permissible objective principle of volition or act-intention in any act-context, unless it consistently generalizes in possible worlds that include our laws of material nature, that is, in worlds in which causality is really possible.

By contrast, the other three formulas of the Categorical Imperative are *material or substantive* procedural moral meta-principles. The Formula of Humanity as an End-in-Itself says that nothing will count as an objective moral principle, and in particular nothing will count as a morally permissible objective principle of volition or act-intention in any act-context, *unless it essentially supports the nondenumerably infinite, absolute, intrinsic, objective value, or dignity, of real human persons by never entailing that they are used as mere means to some end or treated as mere things*. The Formula of Autonomy says that nothing will count as an objective moral principle, and in particular nothing will count as a morally permissible objective principle of volition or act-intention in any act-context, *unless it essentially supports the self-legislating freedom of real human persons*. And finally the Formula of the Realm of Ends says that nothing will count as an objective moral principle, and in particular nothing will count as a morally permissible objective principle of volition or act-intention in any act-context, *unless it essentially supports the self-legislating freedom of real human persons in a universal intersubjective community such that each real human person is considered equally or impartially in the free choices or acts of every other real human person*.

Curiously, Kant says that there are "three ways of representing the principle of morality" (*GMM* 4: 436, underlining added), namely, FUL, FHE, and FRE; but that is clearly just Homer nodding and miscounting, since he has actually provided four formulations in the immediately preceding run of text. So, charitably, what Kant really means is that there are *four* ways of representing "the principle of morality" (underlining added), namely FUL, FHE, FA, and FRE. Now, granting that charitable reading, then precisely how many Categorical Imperatives are there? One or four?

The correct answer is: *both*. This is because *the* Categorical Imperative is most correctly construed as *one* set of *four* lexically-ordered, analytically interderivable, and necessarily equivalent procedural moral meta-principles, one of which (FUL) is also the formal logical, semantic, and normative presupposition for the other three considered as a group (FHE, FA, and FRE). Each of these procedural meta-principles occupies a certain normative-semantic position in the overall moral structure of Existential Kantian Ethics; each plays a certain normative-semantic role, within one and the same larger lexically-ordered, hierarchical moral system of Existential Kantian Ethics's moral principles; and

each differs from the others only in its specific functional normative-semantic nature and in its finegrained intensional content:

> [T]he above [four] ways of representing the [categorical imperative] are at bottom only so many formulae of the very same law, and any one of them unites the other [three] in it. (*GMM* 4: 436)

Here is a directly relevant mathematical analogy. Consider the following statements, T1 to T4, four different ways of thinking about triangles. And, to make the direct relevance of the moral-mathematical analogy even more obvious, let us call the complete set of four statements, *The Triangularity Imperative*:

> T1: As a geometer, you must think that triangulars are triangulars.
> T2: As a geometer, you must think that trilaterals are trilaterals.
> T3: As a geometer, you must think that triangulars are trilaterals.
> T4: As a geometer, you must think that trilaterals are triangulars.

Now statements (T1) through (T4) are all analytically interderivable and necessarily equivalent a priori truths, each of them expressing The Triangularity Imperative, but they are not synonymous. Moreover, T1, as embedding a straight-out identity statement about triangles, is a formal presupposition of T2 to T4. So too, according to Kant and also Existential Kantian Ethics, the Formula of Universal Law (aka the Formula of the Universal Law of Nature), the Formula of Humanity as an End-in-Itself, the Formula of Autonomy, and the Formula of the Realm of Ends are all analytically interderivable and necessarily equivalent a priori moral principles, but they are not synonymous, and the Formula of Universal Law is a formal presupposition for the other three. Just like T1 through T4, each of the several distinct formulations of *the* Categorical Imperative is conceptually or intensionally distinct from all of the other formulations in a semantically finegrained way. Yet at the same time they all belong to a single, multi-termed holistic conceptual network,[109] which, in turn, is fully embedded within one and the same larger hierarchical system of principles, whether moral or mathematical.

What makes this moral-mathematical analogy not merely *directly* relevant but also *deeply* relevant, is the fact that, just as T1 through T4, aka The Triangularity Imperative, is a single set of four analytic truths about how, as a geometer, you must think about triangles, whose subject-matter belongs to the synthetic a priori exact science of geometry, so too the four distinct formulations of *the* Categorical Imperative are all analytic meta-procedural principles about first-order moral principles, whose subject-matter belongs to the synthetic a priori human science (*Geisteswissenschaft*) of morality.

The Kantian theory of moral principles, as I am understanding it from the standpoint of Existential Kantian Ethics, is not only deeply analogous to *mathematics*, as Ross notes: it is also deeply analogous to *logic*. Indeed, I have argued explicitly, in *Rationality and*

Logic and *Cognition, Content, and the A Priori*, that logic and morality are essentially connected.[110] The essential connectedness of logic and morality is particularly salient when we jointly consider contemporary Kantian approaches to philosophical logic and to morality alongside each other. Then it is clear and distinct that there is a significant structural analogy between

> (i) the logico-normative role of *the Formula of Universal Law* in Kant's metaphysics of morals, and
> (ii) the logico-normative role of *the Principle of Non-Contradiction* in Kant's pure general logic.

The classical Principle of Non-Contradiction says that necessarily, no statement is such that both it and its negation are true. Or equivalently, in Kantian terms, since Kant presupposes universal bivalence in pure general logic, the pure general logic version of the Principle of Non-Contradiction says that necessarily, no statement is such that it is both true and false. Hence, according to the classical and Kantian pure general logic versions of the Principle of Non-Contradiction alike, there can be no "truth-value gluts" or "true contradictions." But in view of recent and contemporary work in non-classical logic, especially including dialetheic paraconsistent logic,[111] there is good reason to reject universal bivalence.

In *dialetheic* systems, truth-value gluts or true contradictions are statements that receive both classical truth-values, True and False, on some interpretations, including some theorems of logic. For example, arguably both the Liar Sentence (which asserts its own falsity) [112] and also the Gödel Sentence (which provably asserts its own unprovability)[113] are true contradictions. So dialetheic systems that permit the semantic evaluation of either the Liar Sentence or the Gödel Sentence, allow for the existence of true contradictions. Dialetheic systems, in turn, are a sub-species of *paraconsistent* systems. The defining feature of a paraconsistent system is that it includes an axiom which prevents the valid derivation of every statement whatsoever from any given contradiction, a logical phenomenon which is called "Explosion." So let us call that special axiom a *no-Explosion axiom*. By including a no-Explosion axiom, dialetheic paraconsistent systems constrain the logical powers of contradictions in order to accommodate the possibility of true contradictions within the system, while also preventing the state of global inconsistency or complete logical anarchy or chaos, in which *every* statement is a truth-value glut.

In order to appreciate the full logico-semantic and normative force of Kant's pure general logic, we should interpret the Principle of Non-Contradiction *non-classically and in a Low-Bar, minimal, or nonideal rationally normative way*, as a strictly universal, absolutely necessary, and pure a priori logical meta-principle that lays down a necessary logical constraint on what will count as a true first-order statement in any language or logical system, but also allows for the existence of true contradictions in dialetheic

paraconsistent systems. As I mentioned in chapter 1, my own proposal for this Low-Bar, minimal, nonideal, strictly universal, absolutely necessary, pure a priori, categorically normative logical principle is what I call

> Minimal Non-Contradiction:
>
> Accept as truths in any natural language or logical system only those statements which do not entail that it and all other statements in any or all natural languages or logical systems are both true and false.

Minimal Non-Contradiction, in turn, guarantees what I call "minimal truthful consistency." *Truthful* consistency, as such, means that you must accept as truths in a natural language or logical system only those statements which do not entail that *any* argument in that language or system leads from true premises to false conclusions. By contrast, *minimal* truthful consistency means that you must accept as truths in any natural language or logical system only those statements which do not entail that *every* argument in any or all natural languages or logical systems lead from true premises to false conclusions. This latter notion of course is consistent with holding that *some* arguments in that natural language or logical system lead from true premises to false conclusions, and indeed it is also consistent with holding that some arguments in that natural language or logical system lead from the null set of premises to necessarily false conclusions. If so, then some statements in that natural language or logical system are both true and false, hence are true contradictions. So minimal truthful consistency is consistent with dialetheic paraconsistency. In other words, then, Minimal Non-Contradiction essentially secures minimal truthful consistency, and rules out Explosion. Minimal Non-Contradiction is not a strictly truth-preserving logical principle, and not even a strictly consistency-preserving logical principle—hence it is not a High-Bar, maximal, or ideal standard of rational normativity in logic—but it nevertheless strictly rules out global inconsistency, that is, *logical chaos*. Logical chaos, in turn, is the ultimate result of Explosion: If every statement whatsoever in any or all natural languages or logical systems follows from a contradiction, then the negation of every statement whatsoever also follows from a contradiction, and if every statement whatsoever in any or all natural languages is both true and false, therefore every statement whatsoever in any or all natural languages or logical systems is a truth-value glut or true contradiction.

In the 1980s, Hilary Putnam very plausibly argued that the *negative* version of Minimal Non-Contradiction is the one absolutely indisputable a priori truth:

> I shall consider the weakest possible version of the principle of [non-] contradiction, which I shall call the minimal principle of [non-] contradiction. This is simply the principle that *not every statement is both true and false…* [I]f, indeed, there are no circumstances in which it would be rational to give up our belief that *not every statement is both true and false*, then there is at least one *a priori truth*.[114]

Now Putnam and I would disagree on what the nature of apriority is.[115] As I see it, his view of apriority has been too heavily influenced by Quine's critique of the analytic-synthetic distinction and his deflationary epistemic re-construal of the notion of the a priori. But leaving that disagreement aside, my own contemporary Kantian way of making a somewhat similar point, but even more radically, is to say that Minimal Non-Contradiction *just is* the Categorical Imperative, *insofar as* it inherently governs all logic, cognition, science (whether formal, exact, or natural), and theorizing more generally, as rational human activities, as well as all practical and moral activities.

What, more precisely, is the connection between Minimal Non-Contradiction and the Categorical Imperative? The connection has to do with the crucial notion of *construction*, as it is construed according to Kantian constructivism. According to this construal, a specifically *cognitive* construction is how the human faculty of cognition (including the sub-faculties of understanding, logical reason, sensibility, and imagination) generates empirically meaningful or objectively valid judgments as outputs, given intuitions, concepts, and an actual context as inputs, under innately specified categorically normative objective principles. Thus Minimal Non-Contradiction is an innately specified, strictly universal, absolutely necessary, pure a priori, categorically normative, and immanent structural generative objective meta-principle, specifying low-bar, minimal, or nonideal rationally normative logical standards for the cognitive construction of scientific or more generally theoretical knowledge. Also according to Kantian constructivism, analogously to cognitive construction, a specifically *practical* construction is how a person's faculty of desire produces a morally permissible causally efficacious rational choice or moral judgment—that is, a complete act-intention-implemented-in-a-context—as an output, given desires, practical reasons, and an actual act-context as inputs. Correspondingly then, according to the Existential Kantian Ethics theory of moral principles, The Formula of Universal Law says that necessarily no moral principle, including every candidate principle of volition or act-intention, will count as a legitimate and objective moral principle unless it consistently generalizes over the entire domain of moral principles, and thereby rules out the moral equivalent of Explosion, namely moral contradictions or moral dilemmas all over the place, that is, *moral chaos*:

> Accept as objective moral principles in any or all personal lives or communities only those moral imperatives or ought-claims which do not entail that it and all other moral imperatives in any or all personal lives or communities are both permissible (or obligatory) and impermissible.

So the Formula of Universal Law is, essentially, a minimal objective moral meta-principle of non-contradiction, that is, a higher-order or formal objective moral principle that lays down an absolutely necessary and pure a priori minimal truthful consistency and

generalizability constraint on what will count as a first-order substantive ceteris paribus or material objective moral principle in the system of objective moral principles. The Formula of Universal Law tells us what the specific logical character of any first-order substantive ceteris paribus objective moral principle must be. In this way, the Formula of Universal Law is not a criterion of moral epistemology—hence it is not a super-powered *first-order* substantive objective moral principle. Instead, the Formula of Universal Law is a psychologically generative objective meta-principle for the practical construction of first-order substantive ceteris paribus objective moral principles and moral duties.

Thus the Formula of Universal Law, just like Minimal Non-Contradiction in relation to the cognitive construction of theoretical judgments, is an innately specified, strictly universal, absolutely necessary, pure a priori, categorically normative, and immanent structural generative objective meta-principle, specifying low-bar, minimal, or nonideal rationally normative moral *and* logico-semantic standards for the practical construction of rational choices or moral judgments. Now all construction, as an intentional activity of rational human animals, namely real human persons, is in certain basic respects an activity of practical construction. Hence *all* construction, whether cognitive or otherwise, is inherently constrained and guided by the Formula of Universal Law.

But perhaps the most crucial point here is that the Formula of Universal Law says precisely *nothing* about how we are going to be able to apply it to particular candidates for being first-order substantive ceteris paribus material objective moral principles that are suitable for application in particular actual act-contexts. Therefore, even if it turns out to be *epistemically impossible* to apply the Formula of Universal Law effectively to some prospective first-order substantive ceteris paribus objective moral principles, simply because it turns out that some cases are epistemically indeterminate, *this fact has no bearing on the objective validity of the Formula of Universal Law, as long as it turns out that necessarily, any actually implemented first-order substantive ceteris paribus objective moral principle has the logico-semantic property of consistent generalizability*. The problem of epistemic indeterminacy in applying first-order substantive ceteris paribus objective moral principles belongs to moral or practical anthropology, and *not* to the metaphysics of morals (*GMM* 4: 388-389) (*MM* 6: 217).

More specifically however, and by way of working out the rudiments of a logicizing and moralizing psychosemantics, just how does the Formula of Universal Law (or for that matter, the Formula of the Universal Law of Nature, the Formula of Humanity as an End-in-Itself, the Formula of Autonomy, or the Formula of the Realm of Ends) work as a psychologically generative objective meta-principle for the practical construction of first-order substantive ceteris paribus objective moral principles, and also for the construction of morally obligatory or morally permissible objective principles of volition (namely, moral duties and moral licenses)? Two points are crucial here.

First, insofar as it is applied in actual contexts, the Formula of Universal Law properly operates only on *complete act-intention contents and actual or possible act-worlds.* A

complete act-intention content, according to Existential Kantian Ethics, is a *fully meaningful maxim*. That is, a complete act-intention content is a propositional content which includes essentially non-conceptual contents—the contents of sensory intuitions, desires, and feelings, and the semantic contents of directly referential terms—and also conceptual contents in the form of an imperative indexed to an agent, plus whatever background information about that agent, other agents, or the actual world is required to make that content fully meaningful in that actual act-context. In other words, then, a complete act-intention content is *the content of an-act-intention-implemented-in-an-actual-context*.

Possible worlds for Kant, and also according to my contemporary Kantian view, in turn, are nothing more and nothing less than *complete and mutually logically compatible sets of different conceivable ways the actual manifest natural world could have been*. If you added any other concepts to one of these sets, then a contradiction would be entailed. So in other words, according to my contemporary Kantian view, possible worlds are nothing more and nothing less than *maximal derivability-consistent sets of concepts*.[116] Possible act-worlds, in turn, are possible worlds in which human freedom of the will can occur, that is, possible worlds in which psychological freedom, deep freedom (which Kant calls "transcendental freedom"), and wholehearted autonomy can all occur. Since deep freedom is the source-incompatibilistic spontaneous power of a living, conscious rational agent to cause basic intentional actions in the actual natural and social world,[117] then it follows that every act-world is therefore also a manifest natural and social world in which conscious, intentional causation and rational human animal free agency are not only really possible, but also fully actual, natural, and social facts of life.

Second, insofar as it is applied in actual contexts, the Formula of Universal Law properly evaluates the consistent generalizability of complete act-intention contents or fully meaningful maxims *only over relevantly restricted classes of possible act-worlds*—neither over *all logically possible act-worlds*, nor over *all really or synthetically possible act-worlds*, nor even over *all naturally or nomologically possible act-worlds*.

I am now in a position to propose an Existential Kantian Ethics-inspired and logico-semantically-driven solution for the classical problem of Kantian rigorism—the problem of the apparent over-strictness or apparent overly-extended strictly universal scope of Kantian moral principles. What Kant variously calls (and unfortunately *misnames*, due to his failing to note and heed his own crucial distinction between moral principles and moral duties) the *strict duties*, *perfect duties*, or *ethical duties* are, in fact,

> (i) the first-order substantive ceteris paribus objective moral principle telling us *not to lie and not to make false promises* (in effect, the duty to be sincere or truthful), and
> (ii) the first-order substantive ceteris paribus objective moral principle telling us *to preserve one's own faculty for pure practical reason, not to murder other real persons, and not to commit suicide* (in effect, the negative duty not to harm real persons).

By contrast, what Kant variously calls (and again, unfortunately *misnames*) the *meritorious duties*, *imperfect duties*, or *duties of virtue* are, in fact,

> (iii) the self-regarding first-order substantive ceteris paribus objective moral principle telling us *to pursue individual happiness, develop our own talents, and perfect ourselves* (in effect, the duty telling us to pursue Aristotelian *eudaimonia* in a Kantian sense), and
> (iv) the other-regarding first-order substantive ceteris paribus objective moral principle telling us *to promote the happiness of other real persons, provide benefits for them, and protect them, hence treat them with kindness* (in effect, the positive duty telling us to maximize public utility in a Kantian sense, including providing positive goods for other real persons and also preventing harm to them, but also including an egocentrically-centered, Kantian version of the Aristotelian virtue-obligation to choose and act with kindness towards others).

Granting these distinctions, then the two points I made a few paragraphs above effectively solve the problem of Kantian rigorism by guaranteeing that *no first-order substantive ceteris paribus objective moral principle*—and, in particular, none of the so-called strict, perfect, or ethical duties—*will ever have a strictly universal scope that is more extended than some relevantly restricted class of possible act-worlds*. By the very nature of the logico-semantic evaluation of complete act-intention contents of fully meaningful maxims, then, there simply cannot be a first-order substantive ceteris paribus objective moral principle that is overstrict, overgeneralized, or overextended in its strict universality. Therefore, by the very nature of the logico-semantic evaluation of complete act-intention contents of fully meaningful maxims, the so-called strict, perfect, or ethical duties *simply cannot be applied to any cases to which they do not actually already apply*.

These same points also solve the problem of what might be called *Kantian under-rigorism*, which is that the so-called meritorious duties, imperfect duties, or duties of virtue, seem to be *not strict enough*, in that they do not seem to hold in all cases in which the so-called strict, perfect, or ethical duties also hold. But this problem disappears too, as soon as we realize that the so-called perfect and imperfect duties, namely, the first-order substantive ceteris paribus objective moral principles, *must have exactly the same modal extension or reference*, namely, *they must apply to all and only the members of some relevantly restricted class of possible act-worlds*. They both tell us what we ought to do in all and only such act-worlds, other things being equal. Then both the so-called perfect duties and also the so-called imperfect duties must have exactly the same modal scope, namely the total domain of possible act-worlds, under a ceteris paribus condition.

What then is the significant difference between the so-called perfect and imperfect duties? My proposal is that their significant difference is *not* at the level of modal extension or reference, since they are all first-order substantive ceteris paribus objective moral principles with the same modal scope, but *instead* at the level of modal intension or sense. In other words, the so-called perfect and imperfect duties are just different *modes of*

presentation of the same class of act-worlds, that is, different egocentrically-centered and consciously-grasped aspects, or presented partitions, of the same total domain of possible act worlds.

Then the so-called perfect duties are nothing but first-order substantive ceteris paribus objective moral principles *that seem salient to the moral agent in every possible act-world*, whereas the so-called imperfect duties are nothing but first-order substantive ceteris paribus objective moral principles *that seem salient to the moral agent in all and only the act-worlds in which opportunities for pursuing individual happiness and perfecting oneself, or promoting the happiness of others, positively benefitting them, and preventing harm to them—hence, treating them with kindness—are also salient*. But that difference in mode-of-presentation is perfectly consistent with the fact that all of the so-called perfect and imperfect duties are just first-order substantive ceteris paribus objective moral principles that apply to all and only the members of a relevantly restricted class of possible act-worlds, for any complete act-intention or fully meaningful maxim.

It is crucial to remember here, again, that the question of *how anyone could ever come to know* what the relevant restricted class of possible act-worlds for a given complete act-intention or fully meaningful maxim is, is a completely *separate* moral-anthropological or moral-epistemological question that is simply *irrelevant* to the logico-semantic and normative specific character of the first-order substantive ceteris paribus objective moral principles.

2.4 How to Solve the Problem of Moral Dilemmas

At this point in my discussion, I need to make fully explicit something that I have only briefly sketched above, which is that there are three essentially different logico-semantic types of Kantian moral principles, and that these types should be carefully sorted into a three-levelled hierarchy of principles running from the highest level to the lowest level, as follows:

> LEVEL 1: Absolutely Universal and Objective Moral Meta-Principles, that is, the four analytically equivalent formulations of the Categorical Imperative, knowable by self-evident moral intuition.
> LEVEL 2: Fairly Universal First-Order Substantive Ceteris Paribus Objective Moral Principles, knowable by construction.
> LEVEL 3: Moral Duties and Moral Licenses, that is, the obligatory or merely permissible complete act-intentions or fully meaningful maxims, knowable by moral judgment.

In Existential Kantian Ethics, moral principles inherently guide the intentional acts of real human persons. So this is what I will call *The Hierarchy Interpretation* of the

existential Kantian theory of moral principles, not only in the sense that the principles themselves must be hierarchically sorted in this way, but also in the sense that their psychological realization in the wills of intentional agents carries the same basic hierarchical structure.

****

The Hierarchy of Kantian Moral Principles

LEVEL 1: Absolutely Universal and Objective Moral Meta-Principles, or Analytically Equivalent Formulations of the Categorical Imperative or CI (Scope: They Apply in All Logically and Conceptually Possible Worlds, and Hold for All and Only Finite, Embodied Rational Beings)

+ The Faculty of Pure Practical Reason

which is presupposed by

The Formula of Universal Law [or The Formula of the Universal Law of Nature]

(The Formula of Universal Law corresponds to the logical principle of **Minimal Non-Contradiction**)

which is formally presupposed by

The Formulas of Humanity as an End-in-Itself, of Autonomy, and of the Realm of Ends

which are all presupposed by

LEVEL 2: Fairly Universal First-Order Substantive Ceteris Paribus Objective Moral Principles (Scope: They Apply in All Relevant, Really Possible Act Worlds, and Hold for All and Only Rational Human Beings, Other Things Being Equal)

+ Some Essential Facts about Human Nature

including

1. the need for socialization
2. the need for survival and life
3. the need for individual happiness
4. the need for pleasure and absence of pain

which are all presupposed by

the so-called "perfect duties":	the so-called "imperfect duties":
1. be sincere or truthful	3. perfect yourself
2. do not harm real persons	4. promote others' happiness, prevent harm, be kind

LEVEL 3: Moral Duties and Moral Licenses (Scope: They Apply in the Actual Act-World, i.e. This World, and Hold for All and Only Rational Human Intentional Acts)

+ All the Contingent Facts about Desires and Actual Act-Contexts

which are all presupposed by

Either

Acts with Moral Worth, Done For the Sake of the Categorical Imperative = The Good Will

Or

Acts with Moral Value but not Moral Worth, Merely Conforming to the Categorical Imperative

****

Figure 1. The Hierarchy of Kantian Moral Principles.

In the three-levelled hierarchy of Kantian moral principles, principles at lower levels or types presuppose all the principles at the higher levels or types. Furthermore, the principles occurring at each level are all logico-semantically and normatively equivalent with all the other principles occurring at that level.

More specifically, the various formulations of the absolutely universal and objective Categorical Imperative are all *analytically equivalent with one another across logically and conceptually possible worlds* at LEVEL 1.

In turn, the so-called perfect and imperfect "duties," the fairly universal first-order substantive ceteris paribus objective moral principles, are all *necessarily extensionally equivalent with one another across relevant, really possible act-worlds* at LEVEL 2.

And finally, the moral duties and moral licenses are all *biconditionally equivalent with one another across the actual world* at LEVEL 3.

The modal scope of each of the three levels is also distinct from the others. The absolutely universal and objective moral meta-principles at LEVEL 1, namely, the four analytically equivalent formulations of the Categorical Imperatives, hold for all logically and conceptually possible worlds, or for all and only finite, embodied rational beings, that is, beings possessing an innately-specified capacity, or faculty, of pure practical reason, along with a faculty of desire and a sensibility. My own view is that necessarily, all finite, embodied rational beings are also *minded animals*, and therefore living organisms, in at least a functional sense,[118] but this thesis is not absolutely required for the main point I want to make here.

The main point is that the domain of finite, embodied rational beings comprehended by the moral meta-principles at LEVEL 1 is *not* restricted to human beings, but instead includes any finite, embodied rational beings whatsoever, whether they are human, or artificially-constructed biological systems (as are the "replicants" in Philip K. Dick's famous science-fiction novel, *Do Androids Dream of Electric Sheep*, and in its correspondingly famous film version, directed by Ridley Scott, *Blade Runner*), or alien.

By sharp contrast, the fairly universal first-order substantive ceteris paribus objective moral principles at LEVEL 2 hold for all and only finite, embodied rational *human* beings or real *human* persons, that is, rational beings who are also essentially embodied as living organisms, or animals, in essentially the way we are embodied, and are thereby naturally disposed to pursue and produce happiness.

And finally the moral duties and moral licenses at LEVEL 3, as principles of volition or rational choice, hold for all and only *the actual, context-dependent intentional acts* of rational human beings or real human persons. Here is a compact diagram of the structuralist system of Kantian moral principles I have just been describing:

Now according to Existential Kantian Ethics, as a specifically *nonideal* Kantian ethical theory, the three-levelled hierarchy of Kantian moral principles is governed by two distinct but closely-related level-theoretic structural constraints. Here is the first structural constraint:

> (1) The No-Global-Violation Constraint: In order for a choice or act to be permissible or obligatory in any actual act-context, there cannot be any violation of moral principles of the highest type and at the highest level in the hierarchy of principles. That is, there *cannot be* any violation of any of the analytically equivalent formulations of the Categorical

> Imperative at LEVEL l, *even if there are* violations of first-order substantive ceteris paribus objective moral principles at LEVEL 2.

The No-Global-Violation Constraint strictly forbids violations or inconsistencies of any absolute moral principles at LEVEL 1, which would be global or context-invariant violations or inconsistencies of moral principles, even though the Constraint also permits, in some act-contexts, local or context-sensitive violations or inconsistencies of first-order substantive ceteris paribus objective moral principles at LEVEL 2.

Here is the second structural constraint:

> (2) The Excluded Middle Constraint: If an agent has a moral duty in an actual act-context, then *there is always one and only one moral duty for the agent in that act context*—no matter how difficult it is for the agent herself to discern it—and acting on any *other* conflicting first-order substantive ceteris paribus objective moral principle applying to the agent in that context, is morally impermissible in that context.

It should be obvious how The Excluded Middle Constraint relates to the sharp distinction between *moral principles* and *moral duties*. Even though several different and possibly even really conflicting first-order substantive ceteris paribus objective moral principles might apply to a given intentional agent in a given actual act-context, nevertheless she only ever has *one* moral duty in that context, if indeed she has *any* moral duty at all in that context. The Excluded Middle Constraint is a level-theoretic constraint that guarantees *an empirical moral realism* within the non-platonistic, non-naturalistic (that is, not reductively or scientifically naturalistic) framework of Kantian constructivism in ethics. And it is obviously also intimately related to the fact that the Formula of Universal Law is a formally presupposed, absolutely universal and objective minimal moral meta-principle that imposes global moral consistency on the hierarchical system of moral principles in a *thoroughly nonideal* natural and social world.

I am now in a position to begin to face up to the problem of moral dilemmas on behalf of Existential Kantian Ethics. According to the standard interpretation of the Kantian theory of moral dilemmas, the Kantian ethicist *accepts* that there are *apparent or prima facie* moral dilemmas, but then *denies* that there are *genuine or real* moral dilemmas. The standard interpretation is adequately supported by the (unfortunately, very few) relevant texts in which Kant explicitly discusses moral dilemmas. But the glaring philosophical problem with this standard Kantian theory of moral dilemmas is that it just does not seem to comport with the actual facts about our moral lives on the ground in a thoroughly nonideal natural and social world. That is, everyday moral experience, moral commonsense, and above all, existential insight, all self-evidently tell us *that there are real moral dilemmas*, or at the very least, *that moral dilemmas are really possible*.

So, suppose that we accept either the actuality or the real possibility of moral dilemmas, as some existentially insightful ethicists—for example, Kierkegaard, Sartre, Bernard

Williams, and Martha Nussbaum[119]—would have us do. If so, then the Kantian system of moral principles would apparently lead to contradictions or dilemmas, and "morality totters." The so-called supreme moral principle, the Categorical Imperative, would then be bogus, and pure practical reason would then be nothing but an evolution-generated cognitive self-defense mechanism for deceiving ourselves about *the awful truth*—a morally chaotic world. Tragically, it would be even worse than Hobbes's imaginary state of nature, in which human life, although nasty and brutish, was at least mercifully short. In the world of the bogus and self-contradictory Categorical Imperative, *it would be the end of the world as we morally know it, forever*.

I call this *The Moral Doomsday Scenario*. It is clear and distinctly true that neither the actuality nor the real possibility of The Moral Doomsday Scenario can be permitted by any acceptable version of Kantian ethics. Therefore, in order to avoid both the actuality and the real possibility of moral dilemmas, any acceptable version of Kantian ethics must be able to prove that there *cannot* be such things as real moral dilemmas.

This in turn leads to a meta-dilemma: [120] Either we accept the standard interpretation of the Kantian theory of moral dilemmas, in which case Kantian ethics ends up not being in conformity with everyday moral experience, moral commonsense, and existential insight, by denying the actuality and real possibility of moral dilemmas; or else we accept the claims of moral experience, moral commonsense, and existential insight, in which case we leave Kantian ethics open to the actuality or real possibility of The Moral Doomsday Scenario.

Of course, Kantian ethics is not the only moral theory that has to face up to this sort of meta-dilemma.[121] More generally, for any normative ethical theory grounded on universal principles, either one rejects the existence of moral dilemmas, in which case one's meta-ethics ends up not being in conformity with everyday moral experience, moral commonsense, and existential insight, by denying the actuality and real possibility of moral dilemmas; or else one accepts the claims of moral experience, moral commonsense, and existential insight, in which case one leaves one's meta-ethics open to the actuality or real possibility of a relevant analogue of The Moral Doomsday Scenario.

Now, what should we do? To be sure, every contemporary Kantian ethicist has his or her own take on the problem of moral dilemmas.[122] As a defender of Existential Kantian Ethics, however, I think that it would be a very good thing indeed if there were a fully distinct *third* candidate interpretation of the Kantian theory of moral dilemmas, over and above the standard interpretation and also the interpretation mandated by everyday moral experience, moral commonsense, and existential insight, and which is also importantly different from those of other contemporary Kantian ethicists. This would be a fully distinct third interpretation that is relevantly supported by the Kant texts just like the other interpretations, and independently defensible, but also resolves the meta-dilemma.

What might such a fully distinct and philosophically adequate third interpretation look like? In view of what I have already argued, one point that should immediately strike us is

the obvious structural analogy between moral contradictions or dilemmas in a Kantian structuralist system of moral principles in a thoroughly nonideal natural and social world on the one hand, and alethic contradictions or paradoxes (antinomies, hypercontradictions) in classical or non-classical logical systems on the other. This in turn provides us with a philosophical working clue: What if we thought about moral dilemmas in the Existential Kantian Ethics-based structuralist system of Kantian moral principles in a thoroughly nonideal natural and social world, in parallel with thinking about logical contradictions or paradoxes in non-classical logical systems? And in particular, what if we thought about a formal analogy between the Existential Kantian Ethics-based structuralist system of Kantian moral principles in a thoroughly nonideal world on the one hand, and dialetheic paraconsistent logical systems on the other?

Here is the formal analogy I have in mind. A dialetheic paraconsistent logic explicitly allows for local, logically restricted, or non-Explosive contradictions, while also explicitly ruling out global, logically unrestricted, or Explosive contradictions that lead to logical anarchy or chaos. Moreover, and now also thinking in terms of logic, semantics, and linguistic pragmatics, we can think of local, restricted, or non-Explosive contradictions as being *context-sensitive and systematically variable*, and of global, unrestricted, or Explosive contradictions as being *context–insensitive and systematically invariant*. Correspondingly then, we can hold that a properly-designed Existential Kantian Ethics-based structuralist system of Kantian moral principles in a thoroughly nonideal natural and social world will explicitly *allow for* real local, restricted, context-sensitive and systematically variable moral dilemmas, while also explicitly *ruling out* global, logically unrestricted, context-insensitive and systematically invariant moral dilemmas, which would Explosively entail moral anarchy or chaos—that is, moral contradictions or dilemmas literally all over the place, moral Doomsday. The actuality and real possibility of local moral dilemmas would conform to everyday moral experience, moral common sense, and existential insight; and ruling out global moral dilemmas would also prevent the possibility of The Moral Doomsday Scenario. So we would then have a third interpretation of the Kantian theory of moral dilemmas that prevents the meta-dilemma, and correspondingly we would also have a richer interpretation of the Existential Kantian Ethics-based structuralist system of Kantian moral principles.

Now, granting me also The Hierarchy Interpretation of the Existential Kantian Ethics-based structuralist system of Kantian moral principles as a working hypothesis, I can then develop it in full conjunction with the paraconsistent logic analogy, and thereby spell out—while also, as I mentioned earlier, riffing on Emerson's famous remark about the hobgoblin of small minds—what I call *The No-Foolish-Consistency Interpretation* of the Existential Kantian Ethics-based structuralist system of Kantian moral principles, in five steps, as follows.

First, the relevant Kantian texts adequately support The No-Foolish-Consistency Interpretation:

> Act only on that maxim by which you can at the same time will that it should become a universal law. (*GMM* 4: 421, underlining added)

> A conflict of duties would be a relation between them in which one of them would cancel the other (wholly or in part)…. But since duty and obligation are concepts that express the objective practical necessity of certain actions and two rules opposed to each other cannot be necessary at the same time, if it is a duty to act in accordance with one rule, to act in accordance with the opposite rule is not a duty but even contrary to duty; so a collision of duties and obligations is inconceivable. However, a subject may have, in a rule he prescribes to himself, two grounds of obligation, one or the other of which is not sufficient to put him under obligation, so that one of them is not a duty. (*MM* 6: 224, underlining added)

On the appropriate reading of the texts,

> (i) the phrase "should become a universal law" means that no moral principle can be self-contradictory in a global, logically unrestricted, context-invariant way, which makes the Formula of Universal Law, when understood independently of its real-world application to particular actual act-contexts, the moral equivalent of *Minimal Non-Contradiction*,
> (ii) the phrase "a collision of duties and obligations is inconceivable" means that there cannot be global, unrestricted, context-invariant moral dilemmas, because this would entail that some moral principles are globally, unrestrictedly, and Explosively self-contradictory,
> (iii) the phrase "ground of obligation" means a first-order substantive ceteris paribus objective moral principle, and
> (iv) the phrase "a subject may have, in a rule he prescribes to himself, two grounds of obligation" means that there can still be real local, logically restricted, context-sensitive moral dilemmas between first-order substantive ceteris paribus objective moral principles.

Second, suppose now that these Kantian texts do indeed adequately support The No-Foolish-Consistency Interpretation in the ways I have just indicated. Then we need only to find, within the Existential Kantian Ethics-based moral structuralist framework of moral principes, a proper analogue of a special paraconsistent no-Explosion axiom, that is, a moral meta-principle which explicitly allows for real local, logically restricted, context-sensitive moral dilemmas in a thoroughly nonideal natural and social world, while at the same time explicitly constraining their logico-moral powers so that logico-moral chaos cannot Explosively result from them. Correspondingly, my proposal is that we adopt the following Existential Kantian Ethics-based moral meta-principle, which, as I mentioned above, I call The Lesser Evil Principle:

> Given a real local moral dilemma between first-order substantive ceteris paribus objective moral principles, you ought to choose the first-order substantive ceteris paribus objective moral principle which in that actual act-context is the lesser of the several evils, in the sense

> that acting on it keeps rational faith with the Categorical Imperative to the greatest possible extent.[123]

By virtue of The Lesser Evil Principle, any real violations of or real inconsistencies between first-order substantive ceteris paribus objective moral principles (that is, principles at LEVEL 2 in the hierarchy), are automatically resolved by the moral agent's being required, in any given actual act-context, to choose and act on the first-order substantive ceteris paribus objective moral principle that is the lesser of several evils in that context, in the sense that this is the moral principle which in that actual context keeps rational faith with the Categorical Imperative to the greatest possible extent.

What, more precisely, do I mean by a moral principle's "keeping rational faith with the Categorical Imperative to the greatest possible extent" in a given actual act-context? One thing I mean is that a moral principle MP_1 keeps rational faith with the Categorical Imperative to the greatest possible extent in a given act-context if and only if

> in that act-context MP_1 belongs to a holistic conceptual network which necessarily analytically includes all of the four necessarily equivalent formulations of the Categorical Imperative, as instantiated or implemented in that act-context —hence MP_1 is either analytically entailed by or analytically entails the Categorical Imperative in that act-context—and any other relevant moral principle MP_2 in that act-context is either merely analytically consistent with or else analytically rejected by that holistic network.

And another thing I mean is that a moral principle MP_1 keeps rational faith with the Categorical Imperative to the greatest possible extent in a given act-context if and only if

> (i) in that act-context MP_1 *adequately expresses* the Categorical Imperative, and
> (ii) any other relevant moral principle MP_2 in that act-context is either (iia) merely consistent with the adequate expression of the Categorical Imperative, or else
> (iib) *fails adequately to expres*s the Categorical Imperative.

I am taking these two formulations of the notion of rational-faith-keeping-to-the-greatest-possible-extent, in terms of either *analytic entailment* or *adequate expression*, to be necessarily equivalent.

Third, we explicitly note that insofar as The Lesser Evil Principle and The Excluded Middle Constraint together select a certain first-order substantive ceteris paribus objective moral principle to be the agent's duty in an actual act-context, all sorts of ineluctably actualist or real-world factors will contribute to determining precisely which moral principle is to be chosen, including human desires and causal act-consequences of acts, amongst which will be private utility and public utility. But it would be a serious non sequitur to conclude from this, that the selected principle which is my duty in that context must be an ethically egoistic or act-consequentialist principle, or that it has been selected

for ethically egoistic or act-consequentialist reasons. More generally, it is crucial not to confuse *the phenomenon of context-sensitivity and systematic variability in Existential Kantian Ethics* with *either ethical egoism or act consequentialism*. And it is equally crucial not to fall into an instrumentalist fallacy of thinking that just because every intentional act in this thoroughly nonideal natural and social world *begins with* human desires and causal act-consequences, therefore its guiding principle must be *dependent on* human desires and act-consequences. It should be obvious that there is a close parallel here with Kant's famous warning against falling into *the Empiricist fallacy* of thinking that just because every cognition *begins with* experience, therefore its content or truth must be *derived from* experience (*CPR*: B1).[124]

Fourth, The Lesser Evil Principle also captures a plausible reading of this relevant Kantian text:

> When two such grounds [of obligation] conflict with each other, practical philosophy says, not that the stronger obligation takes precedence … but that the stronger ground of obligation prevails. (*MM* 6: 224, underlining added)

On the appropriate reading of the text,

> (i) the phrase "when two such grounds [of obligation] conflict with each other" means morally unlucky real-world situations in which there is a local, logically restricted, context-sensitive moral dilemma between first-order substantive ceteris paribus objective moral principles, and
> (ii) the phrase "the stronger ground of obligation prevails" means that you ought to choose the first-order principle which in that context is the lesser of the several evils, in the sense that acting on it most keeps rational faith with the Categorical Imperative.

Fifth and finally, The Lesser Evil Principle together with The No-Global-Violation Constraint and The Excluded Middle Constraint, jointly guarantee that in any act-context in which the agent has a moral duty, there will be one and only one first-order substantive ceteris paribus objective moral principle that is her moral duty in that context. This is because The Principle together with the two Constraints collectively rule out any global, logically unrestricted, context-invariant violation or inconsistency of moral principles in the overall structuralist system of principles, *even in cases* of real local, logically restricted, context-sensitive violations or inconsistencies of first-order substantive ceteris paribus objective moral principles, that is, *even in cases* of real local moral contradictions or dilemmas.

One important consequence of The No-Foolish-Consistency Interpretation of the Existential Kantian Ethics-based structuralist system of Kantian moral principles is that The Lesser Evil Principle reinstates, in a non-classical global version, the classical Kantian "*ought* implies *can*" principle, which would otherwise be directly threatened by the

existence of real local moral dilemmas. If The Lesser Evil Principle *does not* hold, then if in some actual act-context I am obligated to choose or do *A* (for example, lie to a murderer because otherwise he will murder an innocent person) and also in that context I am obligated to choose or do not-*A* (that is, not lie to the murderer because it is always wrong to lie), then morally I cannot choose or do either *A* or not-*A* in that act-context—and *ought* does not imply *can*. But if The Lesser Evil Principle *does hold*, together with The No-Global-Violation Constraint and The Excluded Middle Constraint, then I am obligated to choose or do either *A* or not-*A* in that act-context, depending on which is contextually selected to be my duty by the Principle together with the Constraints, and then it follows that either morally I can do *A* (because in that context I am obligated to do *A*) or morally I can do not-*A* (because I am obligated to do not-*A*), but not both—and *ought* implies *can* again.[125]

It is crucial to notice that the process of discovering just *which* of the conflicting first-order substantive ceteris paribus objective moral principles is the lesser of the several evils in that actual act-context is an issue for *moral judgment*, but not an issue for the non-classical skinny logic and the non-classical fat semantics of Kantian moral principles. The fact that it may be extremely or even almost impossibly difficult for a moral agent to discover just *which* principle is the lesser of the several evils in that actual act-context—for example, the poignant case in which you are lucidly asked by your suffering sibling, parent, or life-partner to assist in his or her suicide, as per the real-world case of Stephan and Edith Körner—is entirely irrelevant to the application of The Lesser Evil Principle to that act-context. The Lesser Evil Principle rules out the possibility of any global moral dilemma, and it thereby rules out The Moral Doomsday Scenario. But it does not in any way underestimate or undermine the force and significance either of real local moral dilemmas, or of the very real epistemic difficulties of moral judgment in these actual act-contexts, or indeed of the very real human pathos and tragedy of such situations.

As a consequence of this last point, we rightly feel deep moral pity for Sophie in *Sophie's Choice* when she must choose which of her two children will be killed by the Nazis; and we also rightly feel a deep moral terror when we imagine her awful situation. We thereby fully acknowledge the reality of local moral dilemmas, and also fully acknowledge the corresponding problem of moral judgment in local moral dilemma situations. But it would be no moral comfort whatsoever to Sophie or anyone else if the Moral Doomsday Scenario were true, if the Categorical Imperative turned out to be bogus, if moral chaos were lurking like a post-ethical Godzilla behind the thin façade of everyday etiquette and prudential conduct, and if human morality and rationality went into the abyss. There is a world of difference between introducing *a tragic sense of life*[126] into Kantian ethics on the one hand, and being a moral terrorist like Peter Verkhovensky in Dostoevsky's *The Devils* on the other hand. Fully acknowledging the existence and implications of real *local* moral dilemmas is a deep and ultimately life-affirming and morality-affirming existentialist insight. But asserting or allowing for the existence and

implications of real *global* moral dilemmas is, emphatically on the contrary, tantamount to moral nihilism.[127]

In any case, we can now see that The No-Foolish-Consistency Interpretation of the Kantian theory of moral principles is specifically designed to impose on the total structuralist hierarchy of Existential Kantian Ethics-based moral principles precisely the same sort of invulnerability to global, logically unrestricted, context-invariant moral dilemmas that Alfred Tarski's meta-linguistic solution to the semantic paradoxes (and in particular the Liar sentence) imposes on natural languages and logical systems.[128] Tarski's idea was to stipulate that the logical and semantic predicates, especially including the truth-predicates, for a given language, always belong to its *meta-language*. Then it is impossible to form a Liar sentence in any language. It is only if the hierarchy of languages and meta-languages is collapsed into a single undifferentiated "flatland" language that paradoxes can occur. In a precisely analogous way, in the context of the Existential Kantian Ethics-based structuralist system of moral principles, it is only if all moral principles were treated as if they logically, semantically, and normatively belonged to the same level, that global moral dilemmas could occur. Just to give it a convenient name, I will call this rationally disastrous fallacy of collapsing the levels of the hierarchy of principles, whether in logic or morality, *the flatlander fallacy*.

In the mid-1970s, Saul Kripke noticed that, due to the irreducible presence of indexical terms in natural languages, instances of The Liar Paradox can occur contingently, and also that natural languages can consistently contain their own truth-predicates.[129] Ruth Barcan Marcus then also noticed that moral dilemmas can have essentially the same logical and semantic structure as contingent occurrences of The Liar, even when sets of first-order moral principles are consistent.[130] Correspondingly, according to The No-Foolish-Consistency Interpretation, real context-sensitive and systematically variable conflicts of first-order substantive ceteris paribus objective principles are allowed to occur at the second level of the hierarchy of moral principles. This adequately captures the standpoint of everyday moral experience, moral commonsense, and Existentialist insight. But at the same time, we are rationally enabled *to live with* these real local moral contradictions or dilemmas in the thoroughly nonideal natural and social world, precisely because The Lesser Evil Principle holds, moral Explosion is thereby prevented, and The Moral Doomsday Scenario is thereby ruled out.

Taken together, then, The No-Global-Violation Constraint, The Excluded Middle Constraint, and The Lesser Evil Principle collectively guarantee *the intact good will* of any rational "human, all too human" agent who cognitively constructs, practically constructs, and then wholeheartedly volitionally implements the hierarchical Existential Kantian Ethics-based structuralist system of moral principles in our thoroughly nonideal world. Here we must always remember that, according to Existential Kantian Ethics, as a version of Kantian constructivism, a system of moral principles is *neither* an abstract object in moral platonic heaven *nor* a shifting bundle of merely natural facts. Instead, this Kantian

constructivist system of moral principles is nothing more and nothing less than *a categorically normative immanent structure in a rational human minded animal's or real human person's actual conscious and self-conscious will*; and this immanent structure has both irreducible psychological reality and also robust causal efficacy.

2.5 POLICY OF TRUTH: THE MURDERER-AT-THE-DOOR REVISITED

So much for the adagio movement of this chapter's philosophical symphony. Now for the minuet. More specifically, I want to apply The No-Foolish-Consistency Interpretation of the Existential Kantian Ethics-based structuralist system of moral principles to the classic example of a moral dilemma in Kantian ethics: the notorious murderer-at-the-door case, described in Kant's essay, "On a Supposed Right to Lie from Philanthropy."

As everyone who has ever taken or taught an Introductory Ethics course knows, in this classic and indeed all-too-familiar example, a murderer appears at the door of your house and demands to know whether your friend is inside (*RTL* 8: 425). The context makes it evident that he intends to kill your friend. You can either tell the truth to the murderer and let him kill your friend, or else you can prevent harm to your friend by lying to the murderer. The context again makes it evident that you have no other relevant choices—for example, saying nothing but still somehow fooling the murderer, physically overpowering him, or calling the police in time to stop him.

What makes this example not merely notorious but indeed *notoriously* notorious is that, as everyone who has taken or taught Intro Ethics *also* knows, in "On a Supposed Right to Lie" and elsewhere Kant himself *takes the hard or purist line*, and insists that you must tell the truth to the murderer and let him kill your friend. His rationale is that truth-telling or not-lying is a so-called strict, perfect, or ethical "duty":

> To be truthful (honest) in all declarations is … a sacred command of reason prescribing unconditionally, one not to be restricted by any conveniences…. Every individual … has not only a right but even the strictest duty to truthfulness in statements that he cannot avoid, though he may harm himself or others. (*RTL* 8: 427-428) (See also *LE* 27: 59-60, 254, 257, 444-450, 604-605, 699-702)

In other words, for most contemporary philosophers, Kant comes off as a *moral fanatic* about not lying and truth-telling. For, quite apart from the notorious murderer-at-the-door case and relevantly similar cases, sharply contrary to Kant, almost everyone believes that

> (i) failing "to-tell-the-whole-truth-and-nothing-but-the-truth," and
> (ii) so-called "white lies," that is, according to the *Oxford English Dictionary*, "harmless or trivial lies, especially those told to avoid hurting someone's feelings,"

are both very often *morally permissible*, depending on the act-context. Even so, it remains true that, proceeding here both casuistically and with maximum interpretive charity, Kant *could* be gotten off the philosophical hook. For we *could* note that the complete title of his essay is "On a Supposed Right to Lie from Philanthropic Motives," which can then be taken to say that what he is rejecting is not the moral permissibility of lying *as such*, but only the moral permissibility of lying *from philanthropic motives*. This, of course, is perfectly consistent with claiming that, provided an agent of good will had only non-selfish, non-egoistic, non-hedonistic, non-consequentialist, and thus *non*-philanthropic motives for lying in certain special act-contexts, then there could still be a right to lie from respect for the Categorical Imperative and for the dignity of persons—hence, a right to lie from strictly *moral* motives—in those special act-contexts. And in effect, that is precisely what The No-Foolish-Consistency Interpretation of the Existential Kantian Ethics-based structuralist system of moral principles will ultimately yield as a result.

Nevertheless, the historical and textual evidence for Kant's own personal moral fanaticism and purism about not lying and truth-telling really is just *too* overwhelming. So we must frankly admit that Kant is clearly and even scandalously mistaken on this important point. He has not only forgotten or overlooked the logical, semantic, and normative implications and resources of his own theory of moral principles for solving the problem in a satisfactory way. He has also fallen headlong into the disastrous *flatlander fallacy* of forgetting or overlooking the essentially hierarchical structure of his own system of moral principles and mistakenly treating all moral principles as if they occurred at the very same level. Or in other words, on *this* particular point, I think that Kant *badly screwed up*, and also that the 1980s new wave band *Depeche Mode* were much wiser than Kant:

> Hide what you have to hide
> And tell what you have to tell
> You'll see your problems multiplied
> If you continually decide
> To faithfully pursue
> The policy of truth

Still, as I mentioned before, I also think that by standing straddle-wise on the shoulders of Kantian ethical giants like Kant himself and Ross, we can see a little further than they did. The core of the solution provided by The No-Foolish-Consistency Interpretation, then, is that the murderer-at-the-door case, and all cases relevantly like it, are *real local, logically restricted, or context-sensitive moral dilemmas*, hence real conflicts between first-order substantive ceteris paribus objective moral principles. But at the same time, they do not constitute global, logically unrestricted, or context-invariant violations of the hierachical structuralist system of moral principles—hence *they are never real conflicts of duties, and thus they are never real violations of the Categorical Imperative*.

So let us assume that The No-Foolish-Consistency Interpretation is correct, and also that The No-Global-Violation Constraint, The Excluded Middle Constraint, and The Lesser Evil Principle all hold. What are we to do about the murderer-at-the-door?

In this act-context, each of your two options is an evil. On the one hand, if you tell the truth to the murderer and let him kill your friend, then you have allowed the murderer to harm your friend and have violated the first-order substantive ceteris paribus objective moral principle requiring us to produce positive benefits for others and prevent harm to them. But on the other hand, if you prevent harm to your friend by lying to the murderer, then you have violated the first-order substantive ceteris paribus moral principle requiring us not to lie, other things being equal. So you are between a rock and a hard place.

Nevertheless, *clearly and distinctly*, the lesser evil in this context is preventing harm to your friend by lying to the murderer. It is wrong to lie, other things being equal: but in this context you are morally obligated to lie. Again, it is not just *morally permissible* that you lie in this context: it is *morally required* that you lie in this context.

Why? As I have just said, *lying is wrong, other things being equal*, and that is why this option remains an evil. But the commonsense, everyday moral phenomena of permissibly failing to-tell-the-whole-truth-and-nothing-but-the-truth and of permissible "white lies" self-evidently show that other things are very often *not* equal. Correspondingly, by lying in this context there is no global, logically unrestricted, or context-invariant violation of the Categorical Imperative, since you have chosen to lie *solely for the sake of preventing harm to your friend, thereby stopping the murderer from treating her as a mere means or as a mere thing, like the Nazis treated people, for no good reason at all*. This clearly, distinctly, and most deeply keeps rational faith with the Categorical Imperative in that act-context, by analytically entailing and adequately expressing the Formula of Humanity as an End-in-Itself. And if I am correct that this choice clearly, distinctly, and most deeply keeps rational faith with the Categorical Imperative by analytically entailing and adequately expressing The Formula of Humanity as an End-in-Itself, then since all the formulations of the Categorical Imperative are analytically equivalent, it *also* necessarily follows that it clearly, distinctly, and most deeply keeps rational faith with and thereby adequately expresses the Formula of Universal Law, the Formula of Autonomy, and the Formula of the Realm of Ends *too*.

Moreover, given The Excluded Middle Constraint, it is self-evident that in this act-context, to tell the truth to the murderer would be to accede to, or condone, the murderer's intention to harm and treat the victim as a mere means or as a mere thing, like the Nazis treated people, for no good reason at all, and thus would clearly and deeply violate the Formula of Humanity as an End-in-Itself. So in this act-context, it is self-evidently deeply morally wrong *not* to lie to the murderer. Telling the truth to the murderer in this act-context would be a betrayal of rational faith in the Categorical Imperative, and a failure analytically to entail and express the Categorical Imperative.

Again, it is a moral fact that you *do lie* in this act-context, and that *lying is wrong, ceteris paribus and objectively*. But lying in this context is *not globally wrong*: on the contrary, it is *locally obligatory*, and only *ceteris paribus wrong*. And then there are also all those commonsense, everyday permissible failures to-tell-the-whole-truth-and-nothing-but-the-truth and permissible "white lies." All this, in turn suggests an apt and instructive, if somewhat long-winded, bumper-sticker slogan for Kantian moral guidance in a thoroughly nonideal natural and social world:

> Think Globally, Lie Locally—But *Only* When You Have Overriding Good Reasons To Do So, And *Only For The Sake Of*, Or At Least *Consistently With*, The Categorical Imperative.

And then you must bravely, stoically, and perhaps even life-affirmingly take your hit, and take complete moral responsibility, with no excuses, for your choice to lie, just like Camus's absurd existential hero Sisyphus endlessly pushing his rock up the mountain. Letting the murderer kill your friend by telling the truth, clearly and distinctly, is the massively greater evil in that act-context, and therefore you must choose the lesser evil of preventing murder by lying. Lying is chosen only as the contextually-selected means to the end of preventing murder, which in this context keeps rational faith with the Categorical Imperative to the greatest extent. But you must also *take deep moral responsibility*[131] for your lie. That is the Kant-Sartre Insight.

Here is a further point that needs to be made especially emphatically. In this act-context, *if you could have avoided the lie and still prevented mortal harm to your friend*, say, by saying nothing but still somehow fooling the murderer, or by physically overpowering him, or by calling the police in time to stop him, *then you would have done so*. You do not lie "from philanthropy" or for act consequentialist reasons. On the contrary, *you lie for the sake of the Categorical Imperative, and also for the sake of the human dignity of your friend*, and you also tell that lie *with a tragic sense of life*. So your choice has moral worth, but you also take personal deep moral responsibility for the lie, with no excuses, and you bravely, stoically, and perhaps also life-affirmingly accept whatever awful hit may follow from that complete moral responsibility, whether it be in the form of agent-regret; or in the form of the moral criticism, blame, or punishment applied by to you by others; or even in the form of the sheer cruelty of others towards you. For example, in the tragically all-too-true, post-World War II, Nazi-at-the-door variant on Kant's original case, lying to the Nazi at the door means that you seriously risk *going to the wall for it*, that is, being cruelly murdered too, if your lie is discovered or even seriously suspected.

At the same time, it is also clearly and distinctly true that *anyone else* who actually criticized, blamed, or punished you for this particular lie would be a moral purist fanatic: a *moral martinette*, a *prig*, and *a rule-monger*. –As Kant himself personally seems to have been, in certain respects. Moreover, in the Nazi-at-the-door version, the Nazi killers who would send you to the wall for it, and cruelly murder you, would also be *near-satanically*

evil moral monsters. Of course, the world contains all-too-many moral purist fanatics, and all-too-many near-Satanically evil moral monsters. Our world is a thoroughly nonideal natural and social world, a wheel of Ixion, and vale of tears. It is a once-powerful and arrogant father driven to madness by tragic regret and sorrow, an eyeless man, and a cynical-wise fool, all stumbling together over a storm-blasted heath in Shakespeare's amazing *King Lear*, or in Kurosawa's equally amazing *Ran*. But that isn't *your* fault.

2.6 One Last Thing, by Way of Conclusion

By way of concluding this chapter—the final sonata in this philo-symphony—here is one last thing that I have already anticipated in sections 2.2 and 2.3 above. Corresponding to The Lesser Evil Principle is a further constraint on moral judgment to the effect that a moral agent is required to choose, in any given act-context, what seems to her to be the first-order substantive ceteris paribus objective moral principle that is the lesser of several evils in that context, in the sense that this is the moral principle which in that context seems to the agent to keep faith with the Categorical Imperative to the greatest possible extent. I will call this The Lesser Evil Principle.*

In line with the sharp distinction I noted in section 2.1 between objective moral principles and moral judgments, The Lesser Evil Principle* is not to be confused with The Lesser Evil Principle itself. The moral judgments of "human, all too human" moral agents, perhaps especially including philosophers, are always fallible—so, for example, obviously I could be wrong in my considered moral judgment about which principle to choose in the murderer-at-the-door case, even despite all my rational confidence in that judgment. As a matter of psychological fact, I do not seriously think that I could be wrong about *that* judgment—but of course I *might* be. Nevertheless, errors in human moral judgment are not in and of themselves errors of objective moral principle. We must not confuse the very real difficulties of moral judgment, and the very real human pathos and tragedy of real local moral dilemmas, with the actuality or real possibility of global moral dilemmas. That would be to commit the flatlander fallacy.

Moreover, errors in moral judgment also might *not* be something for which a moral agent can be legitimately criticized, blamed, or punished. More positively put, my Existential Kantian Ethics-based view is that when a moral agent messes up her moral judgment, but still satisfies The Lesser Evil Principle*, then this is not a flaw in her objective moral principles themselves, and not a lapse for which she can be legitimately criticized, blamed, or punished. She cannot be legitimately criticized, blamed, or punished for trying her hardest to select the lesser evil and keep rational faith with the Categorical Imperative. Here we must morally blame, criticize, or even punish *the act* that picked out the greater evil in that context, but we must not morally blame or criticize *the character* of the real person who was wholeheartedly trying to choose the lesser evil.

For example, it seems clearly and distinctly true (to me, anyhow) that the boy in Sartre's story made the morally wrong choice by staying with his mother. He should have joined the Free French Forces. The Nazis were harming and killing literally millions of innocent people and treating them as mere means or as mere things—like pieces of garbage or offal—and they had to be resisted and if possible, stopped. Here we can usefully compare and contrast the boy in Sartre's story with the story of Rick in *Casablanca*: [132] If Rick had stayed with Ilsa and had not joined the Free French Forces, then that would clearly and distinctly have been the morally wrong choice, even though it is equally obvious in the movie that Ilsa is going to be miserable without Rick, and doomed to a loveless marriage with the noble but Karenin-esque Victor Laszlo.[133] In Sartre's case, we can morally criticize the boy's *act*. But I also think that we would be moral purist fanatics to criticize the boy's *character* for his trying wholeheartedly to choose the lesser evil.

In this way, and for the logico-semantically-driven, moral structuralist reasons I have spelled out, it is really possible for us, as defenders of Existential Kantian Ethics, to *live with* real local moral contradictions or dilemmas in this thoroughly nonideal natural and social world. Thereby we can avoid the many small-minded hobgoblins of foolish consistency that are so rightly derided by Emerson, and *also* fully acknowledge Sartre's and others' deep existential insights about these morally tragic situations. Yet at the same time, we are still able to stop well short of the ethical abyss of real global moral dilemmas and The Moral Doomsday Scenario. We can be defenders of full-strength nonideal Kantian ethical theory, emphatically *with* a tragic sense of life, but also *without* becoming moral nihilists. That's what The No-Foolish-Consistency Interpretation is all about.

Chapter 3

NEO-PERSONS AND NON-PERSONS: THE MORALITY OF ABORTION AND INFANTICIDE

> Clearly, it is wrong to kill [arbitrarily chosen] adult human beings. Clearly it is not wrong to end the life of some arbitrarily chosen single human cell. Fetuses seem to be like arbitrarily chosen human cells in some respects and like adult humans in other respects. The problem of the ethics of abortion is the problem of determining the fetal property that settles this moral controversy.[134]

> One reason the question of the morality of infanticide is worth examining is that it seems very difficult to formulate a completely satisfactory liberal position on abortion without coming to grips with the infanticide issue. The problem the liberal encounters is essentially that of specifying a cut-off point which is not arbitrary: at what stage in the development of a human being does it cease to be morally permissible to destroy it?… In the case of abortion a number of events—quickening or viability, for instance—might be taken as cutoff points, and it is easy to overlook the fact that none of these events involves any morally significant change in the developing human. In contrast, if one is going to defend infanticide, one has to get very clear about what makes something a person, what gives something a right to life.[135]

3.1 INTRODUCTION

This chapter, together with the three chapters following it, jointly constitute four successive applications of the nonideal Kantian theory of moral principles—according to The No-Foolish-Consistency Interpretation—that I developed in chapter 2, to several real-world ethical problems and puzzles about *life-and-death issues*. By "life-and-death issues" I mean moral issues concerning

> (i) real human personal life in all its stages from its fetal beginning to its end or termination in death,

(ii) the life-saving or life-harming treatment of other real persons, including non-human minded animals, and
(iii) the meaning or purpose of a real human personal life in full view of the brute fact of our own inevitable and permanent deaths.

The four chapters are therefore obviously closely interlinked in aim and content; but each also provides an independent argument for a set of first-order substantive ceteris paribus objective moral principles about some specific cluster of life-and-death issues.

The present chapter deals with the morality of abortion and infanticide. Chapter 4 is concerned with the morality of our treatment of non-human minded animals, whether sentient and fully minded (like bats or cats) or proto-sentient and "simple minded" (like insects or reptiles). Chapter 5 looks at the morality of saving real human persons' lives, especially including the problematic morality of saving one's own life or others' lives by killing others. And finally, chapter 6 looks at the morality of one's own death in relation to five basic ways it can happen—euthanasia, self-sacrifice, suicide, accidental death, and natural death.

The guiding thread throughout all of this is the thought that the Highest or Supreme Good of real human personal life is the partially or to-some-degree realized life of *principled authenticity*—that is, Kantian-autonomy-with-Kierkegaardian-purity-of-heart—under the Categorical Imperative, in a universal intersubjective community of equally considered (even if not always equally treated) real persons, all living together in the same thoroughly nonideal natural and social world, along with all other morally significant creatures, including sentient animals and non-sentient animals, and other living organisms. Otherwise put, achieving any part or degree of the life of principled authenticity means rationally developing, and then passionately freely choosing and acting upon, as a unified life-project, your own special set of guiding principles, but at the same time wholeheartedly and autonomously following the basic moral principles falling under the Categorical Imperative. And at the same time, it means doing all this while *also* living within a larger holistic ethical framework that is essentially concerned with and naturally driven by life-and-death issues, in the broadest sense of that term, in our thoroughly nonideal natural and social world.

For convenience, I will call this larger holistic ethical framework *The Web of Mortality*.[136] So Existential Kantian Ethics on its applied side puts absolutely universal and objective moral principles fully to work in The Web of Mortality. And in this way, all of applied ethics has an existential Kantian foundation.

3.2 THE PROBLEM OF ABORTION AND INFANTICIDE, AND THE NEO-PERSON THESIS

The familiar—and in certain ways, all-*too*-familiar—problem of the morality of abortion and infanticide, as I am understanding it, is precisely this:

> Is there a property whose non-possession or possession by a fetus or infant necessarily determines the moral permissibility or impermissibility of abortion and/or infanticide? If so, then what is this property?

In the rest of this chapter, I develop and defend a new solution to this familiar problem on behalf of Existential Kantian Ethics, a solution I call *The Neo-Person Thesis*. The Neo-Person Thesis not only directly expresses the basic theoretical commitments of Existential Kantian Ethics and its Kantian nonideal theory of moral principles, but is also importantly distinct from the standard approaches to the morality of abortion and infanticide.

The basic idea behind The Neo-Person Thesis is to ground the morality of abortion and infanticide in the metaphysics of real human persons and in their absolute, nondenumerably infinite, intrinsic, objective moral value, aka their *dignity*—thereby *not* grounding it in *rights*, which, according to my account, actually flow from the dignity of real human persons, and are therefore derivative or secondary moral facts. The Neo-Person Thesis says that since a real human person's life normally begins between 25-32 weeks after conception or fertilization, at which time she becomes a *neo-person*, or "new person," it follows that, *other things being equal, abortion is morally permissible prior to 25-32 weeks following conception or fertilization, but morally impermissible after that time*. It also says that, *other things being equal, abortion or infanticide for non-persons* (a class of creatures that includes human zygotes and fetuses prior to neo-personhood, human fetuses or infants who have been neo-persons but also have permanently lost their neo-personhood by accident or disease, and also human fetuses or infants that never could have been neo-persons due to serious neurobiological abnormalities) *is morally permissible*. In short, then, I am claiming that the metaphysical difference that strictly determines the distinction, other things being equal, between the moral impermissibility and permissibility of abortion and infanticide, is *the difference between neo-persons and non-persons.*

That is the rough-and-ready, or super-simplified, version. More precisely however, what I am saying is that an actualized real human person (for example, you or I or any of the normal adult folks living next door) is literally identical with her third trimester fetus, her neo-person, the new person she was at the front end of her life. This is because, normally between 25 and 32 weeks after conception or fertilization,[137] a human fetus has acquired a constitutively necessary psychological capacity of an actualized real human person, namely the capacity for consciousness—that is, *subjective experience*—even though at that stage of her life and at that time, this human fetus was still only a strongly

potential real human person and not yet an actualized real human person. But since all and only real persons have dignity and are absolutely, nondenumerably infinitely, intrinsically, objectively valuable, then the moral implications of this strong-potentiality-of-a-constitutively-necessary-psychological-capacity are as follows.

First, other things being equal, abortion of any kind is morally permissible prior to the emergence of consciousness or subjective experience, that is, prior to the existence of a neo-person.

And second, other things being equal, abortion of any kind is morally impermissible from that point forwards, except for

> (i) cases of forcible involuntary pregnancy—for example, pregnancy due to rape, and
> (ii) cases in which the mother's life is threatened by the continued existence of the fetus.

These two special exceptions to the impermissibility of abortion after neo-personhood, in turn, depend on three other moral considerations about

> (i) the extent to which we are obligated to prevent harm to innocent real persons,
> (ii) the permissibility of killing in self-defense, even in cases in which the mortal threat is an "innocent attacker," and
> (iii) the inherent badness of one's own "untimely death."

These three considerations will also be separately addressed in more detail in chapters 5 and 6.

Apart from these two special exceptions to the impermissibility of abortion after neo-personhood, there are also some independent sufficient conditions for its permissibility that also suffice for the permissibility of infanticide in a certain special range of cases. Whenever a human being is a non-person—for example, a normal, healthy first or second trimester fetus prior to 25 weeks after conception or fertilization, or alternatively a human fetus at any gestational stage, including infancy, that has such serious neurobiological abnormalities that it never has and never will become even a neo-person, much less a fully actualized real person—then killing it is morally permissible, other things being equal. And some infants are non-persons. Hence infanticide is *sometimes* permissible, *even though infanticide is generally impermissible, other things being equal*. The special exception to the ceteris paribus general impermissibility of human infanticide under conditions of non-personhood, in turn, depends on other moral considerations that apply primarily to minded animals belonging to *other*—that is, non-human—species. These moral considerations will also be addressed separately and in more detail in chapter 4.

One very important upshot of The Neo-Person Thesis, then, is that while the core of the Existential Kantian Ethics-based approach to the morality of abortion and infanticide is grounded in the metaphysics of real human persons (more specifically, in The Minded

Animalism Theory of personhood and personal identity that I develop and defend in *Deep Freedom and Real Persons*, chapters 6-7) it is also *equally* true that significant parts of this approach are also grounded in the morality of harming and saving others, in the morality of our treatment of non-human animals, and in the morality of one's own death. In this way, the moralities of abortion, infanticide, the life-harming or life-saving treatment of others—including non-human animals—and one's own death, all belong coherently, mutually, and ultimately to The Web of Mortality.

That is the basic idea behind The Neo-Person Thesis, drily and formally stated. My dry and formal mode of statement, however, is in no way intended to depreciate or underestimate the intense human pathos and suffering that accompany abortion and infanticide. Nor is it in any way intended to depreciate or underestimate the impassioned sociopolitical debates surrounding them. On the contrary, my mode of statement is dry and formal *precisely* because it fully recognizes all of these facts. Above all, it is intended to emphasize that abortion and infanticide are not atomic, discrete moral concerns, or political litmus tests. Indeed, even despite, and in the face of, the all-too-familiar, morally dogmatic, sanctimonious, and simplistic contemporary American socioculturally- and politically-driven PRO-LIFE vs. PRO-CHOICE dichotomy, abortion and infanticide are almost *immeasurably far* from being simple, one-dimensional moral issues. They belong integrally to The Web of Mortality. My intention, then, is to try to say something clear, distinct, morally illuminating, and non-partisan about several core filaments of The Web of Mortality—filaments that are otherwise too hot for contemporary philosophers and non-philosophers alike, especially in the USA, to handle.

My argument for The Neo-Person Thesis has four parts. First, I state The Neo-Person Thesis and explicate its basic elements (section 3.3). Second, I provide an explicit argument for The Neo-Person Thesis, from Existential Kantian Ethics (section 3.4). Third, I consider and critically respond to the standard approaches to the morality of abortion and infanticide[138] (section 3.5). Fourth and finally, I also consider and critically respond to three possible objections to The Neo-Person Thesis (section 3.6). My overall case for The Neo-Person Thesis thus emerges from the combination of its prima facie plausibility together with its dialectical advantages over the other approaches.

3.3 THE NEO-PERSON THESIS, NEO-PERSONS, AND NON-PERSONS

By *abortion* I mean any intentional act that removes a human fetus from the womb of its natural mother, specifically in order to provide an early termination of the pregnancy, either at the explicit rational request of the mother or at least when the actual or possible rational consent of the mother can be plausibly presumed, whether she has been inseminated by natural or artificial means. This characterization is intended to exclude so-called "spontaneous abortions" (for example, miscarriages) or, more accurately,

unintended abortions, since there can be various kinds of cases in which a fetus is alienated or removed from the womb of its natural mother by accident or without her rational consent or permission.

Within the domain of "abortions" so defined that they are all intended, however, we should also carefully distinguish between

> (i) abortions that detach the fetus from the mother but do not also kill the fetus in the act of detaching it, and
> (ii) abortions that both detach the fetus and also kill it in the act of detaching it.

I will call abortions of the first kind *detachment abortions*, and abortions of the second kind *fatal abortions*.[139] This distinction is deeply important, because the moral permissibility of detachment abortions does not necessarily entail the permissibility of fatal abortions. More specifically, in some cases it will be morally obligatory, other things being equal, to try to preserve the fetus's life during a morally permissible detachment abortion, provided that the medical technology for preserving a detached fetus's life exists. I will come back to this crucial point in section 3.4.

It should also be noted that I am assuming that even if the act of detachment does not kill the fetus, the detached fetus might be killed—whether actively ensuring a fatal result by causally intervening in the fetus's vital processes, or passively ensuring a fatal result by not causally intervening in the fetus's vital processes—before it reaches actualized real personhood. But that is *infanticide*, and not abortion, because by definition according to my account, human infants are all and only the living detached human fetuses prior to actualized real personhood. Similarly, the detached fetus might die by accidental or natural causes before it reaches actualized real personhood. But that is *infant mortality*, and neither abortion nor infanticide.

As I have mentioned several times already, I want to defend The Neo-Person Thesis. When explicitly and fully spelled out, The Neo-Person Thesis has five sub-parts, or sub-clauses, as follows—

> (i) Principle 1: Other things being equal, both detachment abortions and fatal abortions are morally permissible prior to the emergence of fetal consciousness, which normally occurs between 25-32 weeks after conception or fertilization.
> (ii) Principle 2: Other things being equal, after the emergence of fetal consciousness, fatal abortions are morally impermissible except to save the life of the mother, and when a detachment abortion cannot be performed either for purely medical-technological reasons or because it would seriously threaten the life of the mother.
> (iii) Principle 3: Other things being equal, after the emergence of fetal consciousness, detachment abortions are morally permissible only in cases of forced involuntary pregnancy due to, for example, rape, or in order to save the life of the mother.

(iv) Principle 4: At the same time, other things being equal, after the emergence of fetal consciousness, fatal abortions are morally impermissible in cases of forced involuntary pregnancy, and—as stated in Principle 2—morally permissible only in order to save the life of the mother, and when a detachment abortion cannot be performed either for purely medical-technological reasons or because it would seriously threaten the life of the mother.
(v) Principle 5: Other things being equal, infanticide is morally impermissible; nevertheless, when things are not equal, infanticide with respect to an infant X is morally permissible if and only if X is either for some biomedical reason permanently non-conscious, or else X has acquired the capacity for consciousness but for some other biomedical reason X will never become a real person in the natural course of its neurobiological development.

All of these five first-order substantive ceteris paribus objective moral principles obtain *because of* the existence and specific moral character—that is, the "moral status"—of neo-persons, as opposed to the existence and specific moral character or moral status of non-persons. At the same time, the two special exceptions to the ceteris paribus impermissibility of abortion after the emergence of fetal consciousness, and also the special exception to the ceteris paribus impermissibility of infanticide, depend on grounds deriving respectively from the morality of saving and harming others, from the morality of our treatment of non-human animals, and from the morality of "untimely death."

Obviously, everything turns here on the notion of a neo-person. A neo-person is a special kind of real person. More precisely a neo-person is, literally, a *new* real person, that is, a real person *at the front end of her whole life*. As I have worked it out in *Deep Freedom and Real Persons*, section 6.3, the metaphysical analysis of the nature of real persons yields this preliminary, three-part definition of real personhood (where an "*S*-type animal" is simply an animal belonging to to certain species *S*)—

The Preliminary, Three-Part Definition of Real Personhood

Part I. X is a real Frankfurtian person (person$_f$) if and only if X is an S-type animal and X has fully online psychological capacities for

(i) essentially embodied consciousness or essentially embodied subjective experience,
(ii) intentionality or directedness to objects, locations, events (including actions), other minded animals, or oneself, including cognition (that is, sense perception, memory, imagination, and conceptualization), desire-based emotions, and effective first-order desires,
(iii) lower-level or Humean rationality, that is, logical reasoning (including judgment and belief) and instrumental decision-making,
(iv) self-directed or other-directed evaluative emotions (for example, love, hate, fear, shame, guilt, pride, etc.),

> (v) minimal linguistic understanding, that is, either inner or overt expression and communication in any simple or complex sign system or natural language, including ASL, etc., and
> (vi) second-order volitions.
>
> Part II. X is a real Kantian person (person$_k$) if and only if X is a real person$_f$ and also has fully online psychological capacities for
>
> (vii) *higher-level or Kantian rationality*, that is, categorically normative logical rationality and practical rationality, the latter of which also entails a fully online capacity for autonomy (self-legislation) and wholeheartedness, hence a fully online capacity for principled authenticity.
>
> Part III. X is a real person if and only if X is either a real person$_f$ or a real person$_k$; and any other finite, material creature or *entity X* is a non-person.

Now as I also noted in *Deep Freedom and Real Persons*, section 6.2, the beginning of a real person's life, aka its "neo-personhood," is when a given *S*-type animal A manifests the psychological capacity for consciousness and the following counterfactual is also true of A:

> If A *were* to continue the natural course of its neurobiological and psychological development, then A *would* become an actualized real person.

In other words, a neo-person is *not yet* an *actualized* real person, but in fact only a *potential* real person. Nevertheless, a neo-person still really is a special kind of real person and definitely belongs to the total class of real persons by way of a strict identity relation, precisely because the neo-person possesses an essentially embodied innately-specified psychological capacity grounded in a "dedicated" natural neurobiological matrix—that is, a psychological capacity *for consciousness*—that is also the basis of all the other essentially embodied psychological capacities and abilities of actualized real persons. The essentially embodied innate capacity for consciousness, which dynamically emerges only in creatures of suitable neurobiological complexity, is such that it is a true counterfactual fact about each one of them that if that minded *S*-type animal were to go on living in the same way, then it would be an actualized real person.

So the potentiality of a neo-person is what I call the *strong potentiality* of a constitutively necessary psychological capacity—more, specifically, it is a potentiality that supports true counterfactuals about the manifestation of the full range of psychological capacities and abilities that make up the essence or nature of a real person. This is opposed to the *very weak potentiality* of a nomological possibility for being an actualized real person that is possessed, for example, by a given embryo or zygote during the period of

totipotency, provided that it neither splits into twins nor fuses into a chimera. And the strong potentiality of a constitutively necessary psychological capacity is *also* opposed to the *moderately weak potentiality* that is possessed by a given *S*-type animal in the period between the end of totipotency and the emergence of consciousness. In other words, the neo-person has "all the right stuff" for being an actualized real person, whereas a mere embryo or a mere post-totipotency *S*-type animal does *not* have "all the right stuff." Indeed, it is precisely in virtue of this strong potentiality for manifesting the full range of an actualized real person's psychological capacities and abilities that a neo-person constitutes the initial proper part of an actualized real person's whole life. So I am strictly identical with a certain neo-person, my third trimester fetus—the new real person I am at the front end of my whole life *as* a real person.

It follows from all this, that the notion of "a human life" is *a systematically ambiguous notion*. In one sense, "a" human life began when my parents jointly conceived a certain human organism possessing various biological functions, although it was not the life of *an individual human organism* until after the period of totipotency had passed. In another sense, a human life began, after the period of totipotency had passed, in the individual human organism that later became me. But neither of those human organisms were actually *me, myself, I*. So in a third sense, a human life of my own, hence my own real personal human life, began between 25 and 32 weeks after conception.

This line of reasoning enables me to extend the definition of real personhood as follows:

The Extended, Four-Part Definition of Real Personhood

Part I. X is a real Frankfurtian person (person$_f$) if and only if X is an S-type animal and X has fully online psychological capacities for:

(i) essentially embodied consciousness or essentially embodied subjective experience,
(ii) intentionality or directedness to objects, locations, events (including actions), other minded animals, or oneself, including cognition (that is, sense perception, memory, imagination, and conceptualization), desire-based emotions, and effective first-order desires,
(iii) lower-level or Humean rationality, that is, logical reasoning (including judgment and belief) and instrumental decision-making,
(iv) self-directed or other-directed evaluative emotions (for example, love, hate, fear, shame, guilt, pride, etc.),
(v) minimal linguistic understanding, that is, either inner or overt expression and communication in any simple or complex sign system or natural language, including ASL, etc., and
(vi) second-order volitions.

> Part II. X is a real Kantian person (person$_k$) if and only if X is a real person$_f$ and also has fully online psychological capacities for
>
> (vii) higher-level or Kantian rationality, that is, categorically normative logical rationality and practical rationality, the latter of which also entails a fully online capacity for autonomy (self-legislation) and wholeheartedness, hence a fully online capacity for principled authenticity.
>
> Part III. X is a real person if and only if X is either a real person$_f$ or a real person$_k$, otherwise X is a non-person.
>
> Part IV. If X is an actualized real person, then the neo-person of X is also a real person, where the neo-person of X is a given S-type animal A that manifests the psychological capacity for consciousness and the following counterfactual is also true of A:
>
> If A were to continue the natural course of its neurobiological and psychological development, then A would become X.

Given this extended, four-part definition of real personhood, when taken together what what, in *Deep Freedom and Real Persons*, chapters 6-7, I called *The Minded Animalism Theory* of personhood and personal identity, then it follows that the actualized real person X is strictly identical with its corresponding neo-person A.

Before moving on, I need to make one further comment on this extended, four-part definition of real personhood. As we have just seen, for each and every actualized real person, whether this is an actualized real person$_f$ or an actualized real person$_k$, there is a neo-person with whom he or she is strictly identical. So each neo-person who reaches actualized real personhood is strictly identical with a real person$_f$ or a real person$_k$. But of course, and sadly, some neo-persons never in fact reach actualized real personhood. Some sentient human fetuses die—by accident, through disease, or because of abortion, unintended abortion, or infanticide—during the third trimester, at birth, or during early infancy. And some sentient fetuses lose their strong potentiality for actualized real personhood—by accident, or through disease—even though they continue to live on as human individuals. I will call all such neo-persons *doomed neo-persons*.

Do doomed neo-persons count as real persons? Since Part IV is formulated as a conditional whose antecedent specifies actualized real personhood, then doomed neo-persons do not, strictly speaking, count as real persons by the definition of real personhood given in Part III. On the other hand, however, doomed neo-persons are not ruled out by Part III as real persons, until the very moment they either die or lose their strong protentiality, prior to their achieving actualized real personhood. Until then, doomed neo-persons are *candidates-in-good-standing* for being actualized real persons. If neo-persons do in fact manage to become actualized real persons—and I will call all these non-doomed, fortunate neo-persons *successful neo-persons*—then it is a retrospective fact that during

their successful neo-personhood they were real persons at the very beginning of their lives. But for doomed neo-persons, until the very moment of their death or of the loss of their strong potentiality for actualized real personhood, their metaphysical status as real persons is left open-ended in a specially channelled way.

What I mean is that doomed neo-persons, just like successful neo-persons, are *not yet* actualized real persons. But doomed neo-persons are also not *non*-persons at that time. Instead, at that time, they are *prospective* actualized real persons, again just like successful neo-persons. For all neo-persons whatsoever, their metaphysical real personhood status is therefore a *retroactive metaphysical and moral status-fact*, and that real personhood status is triggered only if and at the very moment when the actualized real personhood of that neo-person really happens. Nevertheless, the moral real personhood status of all neo-persons holds just in virtue of their strong potentiality for real personhood, even if they are doomed neo-persons and do not in fact manage to reach actualized real personhood, unlike their luckier counterparts, the successful neo-persons.

This thesis—namely, that the moral real personhood status of all neo-persons holds just in virtue of their strong potentiality for real personhood, *even if* they are unlucky, doomed neo-persons and do not in fact manage to reach actualized real personhood—may seem at first glance very odd. How can a metaphysical or moral status-fact be retroactive? Upon reflection, however, it does seem to capture the truth of the matter accurately. This is for two reasons.

First, there are other distinct, non-question-begging domains in which relevantly similar retroactive status-facts occur as well.

For example, consider the following socially constituted retroactive status-fact. A PhD student who begins publishing before finishing her dissertation, defending it, and graduating with a doctorate, is at the beginning of her professional academic career *only if she actually successfully completes her dissertation, successfully defends it, and actually graduates with a doctorate.* For otherwise she will never qualify for a tenure-track job, or even a contingent faculty job, every one of which (nowadays) standardly requires a PhD, and therefore otherwise she will never have a professional academic career, and so never be at the beginning of that professional academic career. The publications list on her academic *Curriculum Vitae* will start with her first *pre-PhD* publication only if she in fact *receives a PhD.*

Now consider the following *non*-socially-constituted retroactive status facts. The *pre-Socratic* philosophers existed as such only if *Socrates* actually later came into existence and actually became a famous philosopher. The *pre-Cambrian* era existed as such only if the *Cambrian* era actually later came into existence. And so-on.

Therefore the very idea of retroactive status-facts, whether socially-constituted or not, is not in any way ad hoc.

Second, all doomed neo-persons are physically, mentally, causally, teleologically, and morally indistinguishable from successful neo-persons. Even a diseased or sick doomed

neo-person is physically, mentally, causally, teleologically, and morally indistinguishable from a successful neo-person who suffers from exactly the same disease or sickness, and then later recovers and survives. During the period of their neo-personhood, the only significant difference between the doomed neo-persons and the successful neo-persons is a forthcoming fact about their futures, that is, neither a settled fact about their pasts, nor a current fact about their presents. Down the line, this forthcoming fact will of course make a huge difference: the doomed neo-persons will become non-persons at the very moment they die or otherwise lose their strong potentiality for actualized real personhood, and never will be actualized real persons. But, other things being equal, the mere forthcoming fact that A is going to die earlier than B or will lose its strong potentiality for real personhood even though B will not lose that, cannot change A's moral status in relation to B, if by hypothesis they are already otherwise morally indistinguishable. All neo-persons, whether doomed or non-doomed, have the moral status of real persons and must be morally *considered* equally and, in certain crucial respects, also morally *treated* equally, throughout the entire period of their neo-personhood, even though some of them, for whatever brute, contingent reasons, will never in fact become real persons. Similarly, all legitimate PhD candidates must be morally *considered* equally and, in certain crucial respects, also morally *treated* equally, throughout the entire period of their PhD candidacy, even though some of them, for whatever brute, contingent reasons, are in fact not going to finish or successfully defend their PhDs, and, at the very moment of this non-completion or unsuccessful defense, will become non-PhDs. All doomed PhD candidates-in-good-standing are just as legitimate as non-doomed PhD candidates-in-good-standing.

Even granting me all those points, however, it can nevertheless still seem paradoxical that I, who am currently an actualized real person, am also strictly identical with a neo-person. According to The Minded Animalism Theory of personhood and personal identity, my life as a self-identical real person began when I was a neo-person, with the onset of my minded animal capacity for consciousness. Thus my real personal life began at least a year before I became an actualized real person at the stage of my late infancy or early toddlerhood, in the sense that I then actually possessed all the fully online psychological capacities that constituted my non-autonomous Frankfurtian person. But how could an *actualized* real person could be strictly or numerically identical with a merely *potential* real person, even if it is a *strongly* potential one? Doesn't such a difference undermine strict or numerical identity? That can seem paradoxical.

But this appearance of paradox holds only if one fails to recognize the important distinction between

> (i) *the metaphysics of personhood*, which deals with the "What-am-I" question, that is, the question of the nature or essence of a person, and
> (ii) *the metaphysics of personal identity*, which deals with the "Who-am-I"question, that is, the question of the singling-out or individuation of a person.[140]

Real personhood is an essential structure of a certain kind of *thing*. By contrast, real personal identity is an intrinsic spatiotemporal relational property of a whole organismic real personal *life-process* that, at some time and place, eventually reaches actualized real personhood. So real personal identity presupposes actualized real personhood. The individual living organism that eventually reaches actualized real personhood at some time and place therefore must pass through several preliminary stages from the beginning of its complete personal life-process. It starts in the neo-person stage, which is when the human fetus transitions from human non-personhood to the first channelled open-ended stage of his or her career as a real human person, which is, as it were, the *embryo* stage, and carries on until it reaches actualized real personhood in late infancy or early toddlerhood, which is, as it were, the *larva* stage, and then continues on through childhood and teenagerhood, which is, as it were, the *pupa* stage, and then reaches its mature adulthood, which is, as it were, its *imago* stage, and finally proceeds out beyond that stage towards its death, which then closes out its as-it-were *metamorphosis* as a complete, finite, and unique real person.

In other words, and using the familiar biological "embryo-larva-pupa-imago" structure of insect metamorphosis as a simple analogy, my personal identity with a neo-person will seem paradoxical only if one fails to realize that personal identity is a multi-term reflexive, symmetrical, and transitive relation over the several distinct stages of a holistic, spatially situated and temporally irreversible natural complex thermodynamic organismic life-process, and *not* a merely two-term reflexive, symmetrical, and transitive relation over a single static material thing or single static immaterial soul-substance. To be sure, personal identity includes singular material-thing-identity, but it is much more than just that, precisely because I am identical with my complete, finite, and unique life, not merely with a more or less static material thing that conventionally bears my proper name. For example—assuming, of course, that I do not spontaneously combust like the abominable Mr. Crook in Dickens's *Bleak House*, or suffer something equally gruesome that prevents my having an intact corpse when I die—then my corpse will be a more-or-less static material thing that conventionally bears my proper name. But that will not be me: on the contrary it will be nothing but a *mere* material thing, *my lifeless remains*.

A crucial element of The Neo-Person Thesis is the notion of *a psychological capacity*. A psychological capacity is importantly distinct from either a weak potentiality or a strong potentiality in the sense that *all psychological capacities are both weak and strong potentialities, but not all weak or strong potentialities are also psychological capacities.* More specifically, here is how I am construing these two notions—

Weak or Strong Potentiality:

Something X has a weak or strong potentiality for being an F or for doing Y if and only if X is not currently an F, and X cannot currently do Y, but

> either (i) X belongs to an actual natural process P such that X's eventually being F or doing Y as an outcome of P is nomologically possible (weak potentiality),
> or (ii) if X were to continue the natural course of its biological development, then X would become an F or would do Y, other things being equal, because X will eventually develop a natural "epigenetic"[141] mechanism for manifesting F-ness or for causing Y (strong potentiality).
>
> Psychological Capacity:
>
> Something X has a psychological capacity for being an F or for doing Y if and only if X is a living organism, and X has consciousness, and there exists within X a fully online natural epigenetic mechanism for manifesting F-ness or for causing Y that can be triggered or suppressed under a specific set of contextual conditions.

An innately-specified psychological capacity is a psychological *faculty* or *power*, and a learned psychological capacity is a psychological *ability*. In this way, what makes any capacity a specifically psychological capacity *is its being a fully online neurobiological capacity of a living organism with consciousness, consisting in a strong potentiality for manifesting or causing certain special kinds of facts*. In turn, these facts—paradigmatically, facts about rational human cognitive or practical free agency—are all facts about *essentially embodied minds*, aka *minded bodies*, aka *minded animals*.[142]

In any case, something's having a psychological capacity strictly entails its having both a weak and a strong potentiality; nevetheless, something's having either a weak or a strong potentiality does not strictly entail its having a psychological capacity. For example, the packet of biological stuff out of which a living acorn is made has a weak potentiality for being an oak tree, but that packet of biological stuff does not have a strong potentiality for being an oak tree. If I were to squash a living acorn flat with a hammer or my shoe, it would be exactly the same packet of biological stuff, at least in a compositional sense, but it is not true that if that squashed packet continued along in the same way, it would become an oak tree, other things being equal. Strong potentiality requires not merely *an appropriate kind of compositional stuff for the causal outcome that is to be produced*, but also *just the right kind of causally efficacious complex thermodynamic structure for producing that very effect*. In short, strong potentiality entails natural goal-directedness or natural teleology.[143]

By contrast to the mere packet of biological stuff that materially constitutes a living acorn, a living acorn itself has both a weak potentiality for being an oak tree and also a strong potentiality for being an oak tree. Not only is a living acorn's compositional stuff causally appropriate, but also its complex thermodynamic structure is just the right kind for the causally efficacious production of an oak tree. Otherwise put, once living acorns have undergone a certain process of development, then, other things being equal, they do indeed become oak trees. Furthermore, living oak trees have both a weak potentiality and a strong potentiality for having leaves and producing living acorns. This strong potentiality

is triggered in the spring and summer, and suppressed in the fall and winter. But living oak trees did not have this strong potentiality when they were living acorns. Living acorns are not themselves *also*, under appropriate conditions, oak trees. So living acorns, unlike living oak trees, do not have a strong potentiality for having leaves and producing living acorns, although living acorns do have a strong potentiality for being living oak trees. Hence strong potentiality is not a transitive property. Something X can have a strong potentiality for being an F, and all Fs can have a strong potentiality for doing Y, but X nevertheless fails to have a strong potentiality for doing Y. This is because just the right kind of causally efficacious complex thermodynamic teleological structure, for producing *Y* dynamically emerges[144] only diachronically—that is, only over elapsed time—and does not exist in earlier stages of the selfsame life-process.

Consider now a slightly different example, taken from the world of rational human animals. When I was a pre-linguistic toddler, I possessed both a weak and strong potentiality for understanding natural languages, and also—if Chomsky is correct about our knowledge of language, as I believe he (mostly) is[145]—I possessed an innately specified psychological capacity for understanding natural languages, hence I possessed a *language faculty* or *language power*. By contrast, right now I have a weak potentiality for playing the piano but no strong potentiality for it, and certainly no psychological capacity for it, innately-specified or otherwise. No matter what the conditions under which pianos were presented to me right now (as it were, pianos appearing before me at all times of the day and night), I still could not play one. But if I intentionally engaged in a certain goal-directed process of learning and training, then (presumably) I could become a piano player and acquire that psychological capacity as an ability. By means of that intentional process of learning and training, I would have put myself into just the right kind of causally efficacious goal-directed complex thermodynamic organization, or teleological structure, for producing music on the piano. But I do not currently have this ability. I now have a more or less appropriate kind of compositional stuff for playing the piano—for example, I am not made of cotton candy, so that I dissolve when I touch the piano keys. But at the same time, I am not currently thermodynamically and teleologically patterned in just the right way. Nor, I imagine, since (sadly, and no doubt culpably) I have no interest whatsoever in playing the piano, will I ever be.

Finally, let us consider another example, this time taken from the world of human non-minded animals. When the living human organism that was later me—let us call this living organism by the palindromic name *bobhannahbob*, or *bhb* for short, as opposed to me and my corpse, both of which conventionally bear my proper name, 'Robert Alan Hanna', presumably first applied to me at birth—was still an embryo and still within the roughly 14 to 18 day window of totipotency, then he (or at that stage, it) had a weak potentiality for being a minded human animal, and also a weak potentiality for being a real human person. But at that time *bhb* had no strong potentiality for being either a minded human animal or a real human person, since twins and chimeras were still possible for him. Later, after the

period of *bhb*'s totipotency had passed but before the end of the second trimester of his fetal development, then he had a strong potentiality for being a minded human animal and also for being a real human person. But at that time *bhb* had no psychological capacities, since he did not possess a consciousness. Later again, however, after my consciousness had dynamically emerged in *bhb* at the beginning of the third trimester, then I possessed the strong-potentiality-of-a-constitutively-necessary-psychological-capacity for being an actualized real human person, and was therefore a successful neo-person at the very beginning of my own life, and literally or numerically identical with the actualized real person I later became. Of course, all things considered and other things being equal, I am thankful for *bhb*: without him, I would not exist; and all things considered and other things being equal, for real human persons, it is better to be than not to be.[146] Nevertheless, *bhb* was not *me*—just as my corpse (assuming I will have an intact corpse) will be neither *bhb* nor me, but instead only my lifeless remains, which I hereby dub *bob-all-gone*.

3.4 A Five-Step Argument for the Neo-Person Thesis

Against the backdrop of my analysis of real personhood, and the notions of weak potentiality, strong potentiality, and psychological capacity that I have just presented or re-presented in section 3.3, I now want to provide an explicit, step-by-step argument for The Neo-Person Thesis. The argument has five steps.

STEP 1

Real human persons, as specified according to the extended, four-part definition I presented in section 3.3, are also what Kant calls "ends-in-themselves" with absolute, nondenumerably infinite, intrinsic, objective moral value (namely, *dignity*, not *price*):

> Beings the existence of which rests not on our will but on nature, if they are beings without reason, still have only a relative worth, as means, and are therefore called *things*, whereas rational beings are called *persons*, because their nature already marks them out as an end in itself, that is, something that may not be used merely as a means, and hence so far limits choice (and is an object of respect)…. If, then, there is to be a supreme practical principle, and, with respect to the human will, a categorical imperative, it must be one such that, from the representation of what is necessarily an end for everyone because it is an *end in itself*, it constitutes an *objective* principle of the will and thus can serve as a universal practical law. The ground of this principle is: *rational nature exists as an end in itself*. (*GMM* 4: 428-429, italics in the original)

> This principle of humanity, and in general of every rational nature, *as an end in itself* … is the supreme limiting condition of the freedom of action of every human being. (*GMM* 4: 430-431)

> In the realm of ends, everything has either a *price* or a *dignity*. What has a price can be replaced by something else as its *equivalent*; what on the other hand is raised above all price and therefore admits of no equivalent has a dignity. What is related to general human inclinations and needs has a *market price*; that which, even without presupposing a need, conforms with a certain taste, that is, with a delight in the mere purposeless play of our mental powers, has a *fancy price*; but that which constitutes the condition under which alone something can be an end in itself has not merely a relative worth, that is, a price, but an inner worth, that is, *dignity*. (*GMM* 4: 434-435, italics in the original)

Now real human persons, or human ends-in-themselves, or human subjects of dignity, as falling under the Categorical Imperative, are not only *targets* of respect, and thereby must be *considered* respectfully, but also they must always be *sufficiently treated* with respect.[147] Sufficiently treating someone with respect, turn, has three individually necessary, individually insufficient, and jointly sufficient conditions:

> first, someone is sufficiently treated with respect only if she is *not* treated either as a mere means or as mere thing, for example, in the way that Nazis treated people, like a piece of garbage or offal, for no good reason whatsoever,
> second, someone is sufficiently treated with respect only if she is treated in such a way that she can give her actual or possible rational consent to that treatment, and
> third, someone is sufficiently treated with respect only if she is treated *with kindness*—that is, with benevolent attention to her *true* human needs.[148]

These are mutually logically distinct and individually necessary, but still *individually insufficient* conditions for respect. For, despite what may appear at first glance, they are not necessarily equivalent, for two reasons.

First, it is at least minimally really possible for a moral agent to give her actual or possible rational consent to being treated either as a mere means or as a mere thing. Indeed, it is at least minimally really possible that a moral agent could rationally consent *even to becoming someone else's slave or to being killed by that other person*—as an extreme form of self-abasement, self-punishment, self-sacrifice, or sexual self-expression. One real-world example, it seems, is the notorious "German cannibals" case in 2002.[149] But the more general point I am making here is that in all such cases, the moral agent would, of her own free will, *disrespect herself* and therefore be choosing and acting impermissibly.

In *On Liberty*, Mill famously argued that freely willed self-enslavement is impossible.[150] But that is a mistake. Freely choosing self-enslavement is really possible. Self-enslavement is putting oneself in bondage, and thus under a system of harsh external restraints, so essentially equivalent to self-imprisonment—obviously, an extreme form of

putting oneself under a system of harsh external restraints. Both self-enslavement and self-imprisonment are conceptually, metaphysically, and even psychologically coherent, even if, other things being equal, deeply perverse, pragmatically self-stultifying, and morally impermissible. So self-enslavement is not the contrary of freedom.

On the contrary, what I call *Natural Mechanism*, that is, the overwhelming compulsion or manipulation of an agent's choices or acts by inherently deterministic or indeterministic natural processes, hence metaphysical puppethood or robothood, is the contrary of freedom.[151] What is impossible, is to choose freely *while also being* a natural automaton, and this is clearly shown by the soundness of arguments for what is nowadays called *source incompatibilism*[152] in the debate about free will. In rejecting the very ideas of free self-enslavement, Mill confused the concept of *self-stultifying impossibility* with the concept of *freely failing to respect one's own human dignity*. The latter is obviously immoral, but also obviously *not* impossible, since, just like the *ought*, the *ought-not* also implies *can*.

Second, even if someone is *not* being treated as a mere means or as a mere thing, and can *also* give her actual or possible rational consent to some proposed mode of treatment, nevertheless she might still be treated *without* kindness. For example, someone who is living in extreme poverty might receive *just enough food aid* not to starve, and *just enough health care aid* not to die from preventable causes, but also *not enough aid* to be well-fed, healthy, self-supporting, or able to engage in any creative, meaningful, useful, or productive activities. Then she is being *oppressed*,[153] by being condemned to a life of constant neediness and suffering.

The upshot, then, is that a real person is sufficiently treated with respect if and only if

(i) she is not treated either as a mere means or as a mere thing,
(ii) she can give her actual or possible rational consent to that treatment, and
(iii) she is treated with kindness.

In other words, no maxim or fully meaningful act-intention should be chosen or acted upon which entails that real persons are treated either as mere means or as mere things, without their actual or possible rational consent, or with cruelty. To treat a real person without respect, and thus either as a mere means or as a mere thing, without her actual or possible rational consent, or with cruelty, is to harm that real person by violating that real person's dignity. Therefore, other things being equal, it is morally impermissible to harm real persons by violating their dignity; and for the very same reasons, it is also morally obligatory, other things being equal, to prevent or reduce dignity-violating harms to real persons. These are the first-order substantive ceteris paribus objective Kantian moral principles also commonly known as "the negative duty not to harm" and "the positive duty to prevent harm."

Equivalently, a real person is sufficiently treated with respect for her dignity if and only if she is provided with *freedom from oppression*.

In understanding the meaning and implications of these moral principles, however, it is crucial to recognize that not *all* forms of harm to real persons are also violations of their dignity, and therefore *that there are some morally permissible harms*. For example, consider morally permissibly killing a culpable attacker in self-defense, that is, killing a culpable attacker in self-defense by using *minimal lethal force*—the smallest amount and degree of violence that is effective for killing, in a context—when that context is also such that *killing is the only way of stopping that attacker in that context*. Killing someone in self-defense under these two conditions (minimal lethal force in a context, and killing is the only way of stopping that attacker in that context) obviously *harms* that attacker, but *it does not also violate that attacker's dignity*.

Harming someone is doing something bad *to* them, or doing something that is bad *for* them. Other things being equal, killing a real human person obviously harms that real person. This is because, *other things being equal, one's own death*—and by "one's own death," I mean a real person's process of dying, terminated, full stop, by the permanent state of that person's being dead—*is a bad thing for the one who dies,* as I will argue in chapter 6 below. In that chapter, this kind of death is what I call an *untimely death*. By contrast, *morally permissible voluntary euthanasia*—that is, mercy-killing someone for the sake of preventing or stopping what I call *the personhood-destroying suffering* of that real person, not only with the actual or possible rational *consent* of the sufferer, but also in response to the actual or possible rational *request* of the sufferer, as a poignant, solemn act of kindness to that suffering person—would arguably be an exceptional case in which things were not equal.[154] Such a death would be what in chapter 6 I call a *timely death*, and an intrinsically good thing for the sufferer herself.

As I say, all of these important points, especially including the notion of personhood-destroying suffering, will be discussed in detail in chapter 6 below. But for the time being, the crucial thing is just that each of the two cases of

(i) morally permissible killing in self-defense, and
(ii) morally permissible voluntary euthanasia,

clearly shows that not every killing morally disrespects the real person who is killed. By the same token, some harms to real persons are *morally permissible harms*; and all and only those harms that are also violations of the dignity of real persons are *morally impermissible harms*, other things being equal.

STEP 2

A successful human neo-person is strictly and numerically identical to an actualized real human person at the very beginning of his or her life. This is because, after consciousness—namely, subjective experience—has emerged at the beginning of the third

trimester, then a successful human neo-person possesses the strong-potentiality-of-a-constitutively-necessary-psychological-capacity for being the real human person that she actually becomes. It is morally impermissible to harm an actualized real person by violating her dignity, other things being equal. Therefore, since the successful neo-person of that actualized real person is identically the same real person, although at the front end of her life, it is morally impermissible to harm that neo-person by violating her dignity, other things being equal.

What about human neo-persons who do not in fact become actualized real human persons, that is, what about *doomed* human neo-persons? As I argued above, during the period of their neo-personhood, all doomed neo-persons are physically, mentally, causally, teleologically, and morally indistinguishable from *successful* neo-persons, that is, from those more fortunate neo-persons who are literally identical with actualized real persons. Otherwise put, as long as doomed neo-persons are still alive and still have a strong potentiality for real personhood, then there is nothing whatsoever to distinguish them morally from successful neo-persons. And this is because doomed neo-persons differ from successful neo-persons only in the mere metaphysical fact that later they fail to become actualized real persons. So, as long as the doomed neo-persons are still alive and possessed of a strong potentiality for real personhood, then they are morally indiscriminable from the surviving, successful neo-persons. The metaphysical fact that the doomed neo-persons will not make it to actualized real personhood is just a *forthcoming fact* about their future lives, not *a settled or current fact* about their past or present lives respectively. But *since choice and action always actually happen in the present and always arise from the past*, and *since they never directly involve the future except by goal or intention*, then the metaphysical forthcoming fact of a doomed neo-person's death or loss of strong potentiality for real personhood is *not a moral fact*, and is *always morally irrelevant*, other things being equal. Therefore, it is morally impermissible to harm any neo-person, whether successful or doomed, by violating its dignity, whether this dignity flows from its being a real person (namely, from its being a successful neo-person) or from its merely being a candidate-in-good-standing for being a real person (namely, from its being a doomed neo-person), other things being equal.

Now, other things being equal, an abortion performed on the mother of a neo-person would cause harm to that neo-person by killing it—or at the very least, by being a mortal threat to it—and also violate its dignity by treating it either as mere means or as a mere thing, or by mortally threatening it without its actual or possible rational consent. So, other things being equal, it is morally impermissible to abort a neo-person, whether by means of a detachment abortion or a fatal abortion. Nevertheless, other things being equal, at any time prior to neo-personhood, then either a detachment abortion or a fatal abortion may permissibly be performed, assuming the actual or possible rational consent of the mother.

STEP 3

Other things being equal, whenever a human fetus or human infant is a human non-person—that is, a human animal that is neither a neo-person (whether a successful neo-person or a doomed neo-person) nor an actualized real human person—then either a detachment abortion or a fatal abortion may morally permissibly be performed, or an infanticide may be carried out, assuming the actual or possible rational consent of the mother in the case of a fetus, or of any other relevant primary moral guardian(s) in the case of an infant. This is because, other things being equal, non-persons are neither subjects of dignity nor targets of respect, and do not belong to the universal intersubjective community of equally considered real persons or rational animals, The Realm of Ends. More specifically, no non-persons are neo-persons, and no neo-persons are non-persons. For example, both the anencephalic infant Baby Theresa and also Terry Schiavo after her catastrophic heart attack were human non-persons, but not neo-persons. This is because, in those actual contexts,

> either (i) an individual human animal lacked even the weak potentiality to be a real human person by lacking a higher brain and the natural neurobiological matrix of consciousness, and thereby also lacked the strong potentiality to be an actualized real human person (for example, the anencephalic infant Baby Theresa),
> or (ii) an individual human animal had previously been an actualized real person who, unfortunately, became permanently unconscious, thereby dying, even though a different successor non-sentient, non-conscious animal conventionally bearing the same proper name still lived on (for example, Terry Schiavo after her catastrophic heart attack).

Just as a non-person fetus may be morally permissibly killed, other things being equal, so too any infant, toddler, adolescent, or adult human that is a non-person may be permissibly killed, other things being equal.

To be sure, other things are not always equal. Hence there are at least three kinds of special cases in which it is morally impermissible to kill or even saliently harm infants, toddlers, adolescents, or adult humans who are, currently and strictly speaking, non-persons.

First, there are cases of what Jeff McMahan calls *post-personhood*, and the moral protection of such non-persons is grounded on the metaphysical and moral fact *that they have previously been real persons*, and then permanently lost their real personhood through disease, injury, or mental illness, but still retain their sentience or minded animality.

Second, there are cases of what I will call *remediable non-personhood*, and the moral protection of such non-persons is grounded on the metaphysical and moral fact that although a human being has lost his or her real personhood through disease, injury, or mental illness, *nevertheless it remains medically (that is, biologically-technologically)*

possible for them to recover from their current unfortunate state of non-personhood and become real persons again.

Third and finally, there are cases of what I call *associate membership in the Realm of Ends,* [155] and the moral protection of such non-persons is grounded on the fact that some actual real persons have individually or collectively resolved to treat those non-persons *as if* they were real persons.

So some human beings, even despite their currently and strictly speaking being non-persons, nevertheless have high moral status because they are

> either (i) *retrospective and still sentient* subjects of dignity (post-personhood cases),
> or (ii) *prospective* subjects of dignity (remediable non-personhood cases),
> or (iii) *conventional* subjects of dignity (associate membership in the Realm of Ends cases).

STEP 4

There are two special exceptions to the moral impermissibility of aborting neo-persons.

The first special exception is that, other things being equal, it is morally permissible to perform detachment abortions in cases of forcible involuntary pregnancy, for example, pregnancy due to rape. This is because the pregnancy has occurred by means of *coercion* (by which, in this context, I mean: *using people as a mere means, specifically by employing violence or the threat of violence*[156]) and without the actual or possible rational consent of the mother. In other words, the pregnancy has been *forced or imposed* on the mother by someone else. And by having this pregnancy forced or imposed on her, she has been treated without respect and also without her rational consent, hence she has been harmed through a violation of her dignity. So it is not morally required that she herself provide what the innocent neo-person needs in order to survive, other things being equal.

It is crucial to recognize, however, that this would remain equally true if she had been coerced instead to provide what another innocent *actualized real person* needs in order to survive, for example, her bone marrow, one of her kidneys, or the shared use of her internal organs for nine months—as per Judith Jarvis Thomson's famous thought-experiment of the Violinist who has been life-savingly attached to your internal organs without your permission while you are asleep.

By sharp contrast, however, if, other things being equal, the mother of the fetus—corresponding, in Thomson's thought-experiment, to the "attachee" or "host" of the attached Violinist—had been morally required merely to wade into a shallow pond and get her nice clothes dirty in order to save a drowning child, then the specific moral character of the case would correspondingly be sharply different. Indeed and more generally, as we shall see in section 5.3 below, such life-saving acts as wading into a shallow pond in order to save a drowning child are arguably not only morally permissible, but also morally obligatory, for an agent in certain actual contexts, other things being equal, provided that

> (i) the agent is the closest one to the mortally threatened innocent real person,
> (ii) the agent is the only one who can save the mortally threatened other innocent actualized real person in that context,
> (iii) the act of saving the child costs the agent nothing of moral significance, even though the agent does indeed sacrifice something of non-negligible moral value, and also
> (iv) the agent is not required to iterate that small sacrifice to the point at which it undermines her obligatory life-project, other things being equal, of developing her abilities and perfecting herself.

At the same time, however, other things being equal, it is also not morally permissible for the mother to insist on a fatal abortion in cases of forced involuntary pregnancy. This insistence on fatality would be killing the neo-person without also being able to assume the neo-person's possible rational consent, much less the neo-person's actual rational consent, since, obviously, the neo-person is incapable of consenting or even reasoning during that period of her/his life. Hence it would constitute treating the neo-person without respect, and would thereby harm the neo-person by violating her/his dignity. For the mother to insist on the fatal abortion of a neo-person in the case of forced involuntary pregnancy, would be morally equivalent to the following sort of case:

> The same mother permissibly refuses to provide a kidney for some unfortunate, innocent, actualized real person who needs a healthy kidney in order to live, and shows up at her front door one day in order to ask for it specifically from her, and then, after the refusal, that same mother proceeds to *strangle* the unfortunate, innocent, healthy-kidney-needer, in order to ensure that s/he dies *right then-and-there*.

Significantly, the issue of central relevance here is not the *morality* of killing per se, since killing can happen in the morally permissible case of detachment abortions and in the morally impermissible case of fatal abortions, alike. It is instead the *modality* of killing that is of central relevance. A detachment abortion is only ever a *contingent* killing—for if the fetus happens to survive as an infant, then that would be perfectly consistent with the intentional aim of a detachment abortion—whereas a fatal abortion is a *necessary* killing, in the sense that killing the fetus is an intrinsic part of its intentional aim. To insist on a fatal abortion when only a detachment abortion is permissible is therefore to kill the fetus "with extreme prejudice," to borrow a vivid phrase from Francis Ford Coppola's stark and uncompromising 1979 film *Apocalypse Now*, or as Thomson puts it more soberly, "unjustly,"[157] other things being equal. Similarly, other things being equal, refusing to provide a kidney for an unfortunate innocent healthy-kidney-needer is morally permissible. But ensuring the death of that unfortunate innocent healthy-kidney-needer right then and there, by (say) strangling her/him, is morally impermissible, other things being equal, precisely because it kills the innocent healthy-kidney-needer "with extreme prejudice" and

"unjustly." That is, it treats the other real person either as a mere means or as a mere thing, without her/his actual or possible rational consent, and with cruelty, hence without respect, and thereby impermissibly harms that real person by violating her/his dignity.

One deeply important consequence of this point is the following point. Assuming that the medical technology for preserving a detached fetus's life exists, then according to the view I am defending here, *other things being equal, it is morally obligatory to try to preserve a neo-person's life during a morally permissible detachment abortion.* To do otherwise would be to treat the neo-person either as a mere means or as a mere thing—as something entirely disposable like a piece of garbage or offal—or without her actual or possible rational consent, and with cruelty, and thereby impermissibly to harm that neo-person by violating her dignity.

The second special exception to the moral impermissibility of abortion is that it is morally permissible, other things being equal, to perform fatal abortions if

> (i) the continued existence of the neo-person threatens the life of the mother, and
> (ii) a detachment abortion cannot be performed either for purely medical-technological reasons or because it would seriously threaten the life of the mother.

The reason why fatal abortions are permissible, other things being equal, when the mother's life is threatened and a detachment abortion cannot be performed, is the same reason why, other things being equal, it is morally permissible to kill another real person—whether a neo-person or an actualized real person—in self-defense, even if that neo-person or actualized real person is entirely innocent of any wicked intention or act, provided that

> (i) killing is the *only* way one can protect oneself from being mortally threatened in that context, and
> (ii) only *minimal* lethal force is used.

Following Thomson's terminology in another important paper, I will call any mortally threatening innocent person of this kind an *innocent attacker*.[158]

So what I am asserting, by implication, is that it is morally permissible to abort a neo-person who is an innocent attacker, other things being equal. This is because

> (i) at least the *possible* rational consent of the innocent attacker can be assumed in such cases,
> (ii) the innocent attacker is not being treated either as a mere means or as a mere thing, and
> (iii) the innocent attacker is also being treated with kindness—insofar as, when the attacked person fends off the innocent attacker, she intends no cruelty whatsoever towards him.

So in these ways the innocent attacker is being sufficiently treated with respect, and not being harmed by a violation of his dignity, other things being equal.

What is the criterion of *possible rational consent*? The basic idea, as I am construing it, is that if any higher-level or Kantian real human person were placed behind a Rawlsian "veil of ignorance,"[159] which procedurally screens out *uniquely self-referring personal identity details* from that person's own cognitive and practical point of view, and temporarily ensures a suitable reflective disinterestedness and distance from her actual "human, all too human" condition, then, for that higher-level or Kantian real human person, the moral permissibility of self-defense would still hold, *no matter who actually turns out to be the self-defender and no matter who actually turns out to be the innocent attacker*.

For example, consider a scenario in which I am a bicyclist and involved in a two-bicycle accident with another bicyclist, previously unknown to me (so: s/he is specifically *no*t a loved one, a close friend, or someone else I have explicitly or implicitly promised to aid or protect), that is no one's fault—for example, a sudden heavy gust of wind blows me and the other cyclist into one another. But unfortunately the accident happens on a busy street, and now the other cylist is lying unconscious on top of me, while suddenly a large Sport Utility Vehicle (SUV), being driven by a reckless (or drugged-up, or drunk, or texting) college student, is barrelling directly towards both of us at high speed and is just a few yards away, unable either to stop in time or swerve so as to miss both of us. As it so happens, then, absolutely the *only* way I can save myself from being run over by the SUV is to push the unconscious other cyclist towards the speeding SUV, and roll sideways. The unconscious other cyclist is an innocent attacker in this case, and I hold that it would be morally permissible for me to kill him in the way I have described, other things being equal.

The rationale is this. I am morally required, other things being equal, to provide benefits for real human persons, and also to prevent harm to them, *including* myself. Moreover, other things being equal, my untimely death is a bad and harmful thing for me. Also I am morally required, other things being equal, to pursue my own self-perfecting projects, which obviously will not be possible if I am dead. So self-defense is at the very least morally permissible, other things being equal, and is a first-order substantive ceteris paribus objective duty to myself. In this case, I am not treating the innocent attacker either as a mere means or as a mere thing, or with cruelty, and harming them by violating their dignity as a person—there is nothing "personal" in my pushing them off me in that way, thereby killing them. Indeed, if there were any other possible way I could push them off me, save myself, and *also* save their life, then I would do so. Nor am I being unkind specifically to *them*: I intend no cruelty whatsoever towards them. Moreover, I would also give my counterfactual rational consent to a scenario in *which I am killed in exactly the same way*, in a slightly different possible act-world in which our personal identities were switched, and unluckily *I was the unconscious cyclist*, and s/*he was the conscious cyclist accidentally pinned underneath me*. Therefore, in the actual world the unconscious cyclist's possible rational consent can be assumed, other things being equal, and I am also sufficiently treating them with respect and not violating their dignity, other things being equal—even though, obviously, I am seriously harming them by killing them.

On the other hand, however, it must be emphasized that I am certainly *not obligated* to kill an innocent attacker, and therefore its being morally permissible on the grounds of self-defense, other things being equal, does not morally rule out my choosing self-sacrifice. This can be shown by considering a minor variant on the original case I imagined, in which it is still possible for me to push the unconscious other cyclist out of the way of the speeding SUV, but, unfortunately, *only in such a way as to guarantee that I will be run over by the speeding SUV instead*. Otherwise, s/he will be run over and undoubtedly killed if I save myself by wriggling out from under them and rolling out of the way. In such a possible act-world, I *might* choose to save the unconscious other cyclist. If so, then in that act-world I would be a *moral hero* at the cost of my own death. But this moral heroism would be *supererogatory*—that is, over and above what is obligatory. For according to Existential Kantian Ethics, I am not morally *required* to be a moral hero and sacrifice myself, other things being equal; rather I am only morally *permitted* to be a moral hero and sacrifice myself, other things being equal.[160]

Moral heroism for "human, all-too-human" rational agents like us, our "sinner-sainthood," is a High-Bar normative *ideal*, that we are obligated *to pursue* in order to have meaningful lives, not a Low-Bar normative *requirement* that we are obligated *to choose and do* in order to treat everyone, including ourselves, with sufficient respect for their human dignity.

This is another important way in which Existential Kantian Ethics is sharply different from act consequentialism, which would strictly obligate me to sacrifice my life, if the unconscious other cyclist were someone who could bring significant benefits to other people: for example, if s/he were a rich philanthropist, a great surgeon, or a great concert violinist. Indeed, act consequentialism would even obligate me to sacrifice my life if the unconscious other cyclist were merely a *moderately well-off* philanthropist, *a fairly good* surgeon, or *a pretty good* concert violinist. Presumably and realistically, whatever shallow happiness benefits I could bring to other people as an independent philosopher would still, on balance, be *greatly* less than those the moderately well-off philanthropist, the fairly good surgeon, and the pretty good concernt violinist would each bring to people by surviving. In fact, even making the very optimistic and no doubt unrealistic assumption that whatever shallow happiness benefits I could bring to other people as an independent philosopher are, on balance, only a *little* less than those a moderately well-off philanthropist, a fairly good surgeon, or a pretty good concert violinist would bring to people by surviving, neveretheless act consequentialism would still obligate me to sacifice myself. But none of this self-sacrificing activity is required by Existential Kantian Ethics.

This is an especially telling point, because it is often falsely assumed that all Kantian approaches to ethics are highly morally strenuous and even morally over-demanding. Perhaps it is true that *some* Kantian approaches to ethics *are* morally over-demanding. But Existential Kantian Ethics is not. On the contrary, Existential Kantian Ethics's level of moral demandingness, while indeed pretty high, and thereby adequately expressive of the

most important ideals of our rational human moral nature, is neither *under*-demanding nor *over*-demanding, but instead *just demanding enough*.

On the other hand, then, and ironically, by the same token, act consequentialism is *under-demanding*. This is because, in cases in which the unconscious cyclist is just an ordinary person, whose abilities to bring about shallow happiness benefits are even less prodigious than my already, realistically, very meagre shallow-happiness-producing abilities as an independent philosopher, and s/he is most certainly not a philanthropist, surgeon, concert violinist, etc., then act consequentialism would make it morally *impermissible* for me to sacrifice myself for their sake. But if Existential Kantian Ethics is right, then moral heroism, or "sinner-sainthood," is not only morally permissible but even *morally a great thing, no matter what the consequences*. Thus moral heroism or sinner-sainhood is not only fully morally permissible but highly morally laudable, even if supererogatory.

In this respect, as in others,[161] *act consequentialism seriously disenchants the moral world.* What would be the point of my merely living on and on and on, wallowing in my shallow happiness, like Mill's satisfied fool or satisfied pig, if the very possibility of my acting on those highest or supreme values that essentially make real human personal life worth living, is morally ruled out of court? In a sufficiently enchanted moral world, it has to be morally possible for me *to choose to be Mill's "Socrates dissatisfied*." Or at any rate, in a sufficiently enchanted moral world, it has to be morally possible for me, like Plato's Socrates, *to choose my not fleeing Athens, my acceptance of the death sentence imposed on me by the Tyrants, and my drinking the hemlock*.

Now back to abortion. For all the same basic reasons and with all the same basic provisos governing the unconscious other cylist case, therefore, on the strength of that reasoning, whenever a neo-person is an innocent attacker, then a fatal abortion is morally permissible, other things being equal, provided that

(i) killing is the only way the life of the mother can be saved in that context, and
(ii) only minimal lethal force is used.

Otherwise, only a detachment abortion is morally permissible, other things being equal. Moreover, just as in the case of morally permissible detachment abortions when the pregnancy is forced and involuntary, provided that the medical technology for preserving a detached fetus's life exists, other things being equal, it is also morally obligatory to try to preserve a neo-person's life during such an abortion.

STEP 5

Finally, there is a special exception to the other-things-being-equal moral permissibility of aborting or carrying out infanticide on human non-persons, as I indirectly

indicated above. Under certain conditions—broadly speaking, the necessary and sufficient conditions governing the existence and specific character of a normative convention[162]—human or non-human non-persons can be temporarily or permanently treated *as if* they were real human persons falling under under the protection of the Categorical Imperative. They thereby gain an "associate membership in The Realm of Ends," whereby they are *secondary* subjects of dignity and *secondary* targets of respect, and thus *extrinsically* receive a temporary or permanent *right-to-life*.

By a "right,"[163] I mean a subject's moral demand on others to let her choose something, do something, or continue being something (aka "liberty rights"), or to provide her with some good or with access to some good (aka "claim rights"). This moral demand can be

> either (i) unalienable, which is to say that it cannot be removed under any conditions, or else (ii) forfeitable, which is to say that it can be removed under certain conditions.

Correspondingly, by "the right-to-life" I mean

> a subject's unalienable moral demand on others to let her continue being alive, that is, a subject's unalienable moral demand not to be impermissibly actively or passively killed by others.

As unalienable, obviously, the right-to-life is not a forfeitable right of any sort. Nor, however, is it a strict right-not-to-be-killed. For if it were a strict right-not-to-be-killed, then it would implausibly prevent morally permissible killing of any sort, for example, during wartime, in self-defense, and other special cases, especially including Trolley Problem-type cases (see chapter 5 below), in which a few people are permissibly killed in order to save many other people.

The crucial point here is that temporary or permanent possession of a right-to-life by secondary subjects of dignity and secondary targets of respect is in sharp contrast to the possession of dignity by *primary* subjects of dignity and *primary* targets of respect—namely, all real persons, including all actualized real human persons, and all neo-persons. The moral status of associate membership in The Realm of Ends is inherently contingent and extrinsic, precisely because it is conventional, although it also remains normatively and morally binding, precisely to the extent that some primary subjects of dignity and primary targets of respect are prepared to offer some good reasons for this moral convention, to stand behind it, and to provide moral censure of those who violate it.

Now a necessary condition for something's being a secondary subject of dignity and a secondary target of respect is *that it have a morally valuable life of its own*, which in turn implies that it must *at least* be an individual living organism, since this is a constitutively necessary condition of being a subject of dignity and a target of respect. Insofar as associate membership in The Realm of Ends has actually been extended to some non-persons that are alive and human, even if they are neither minimally sentient—or what, in chapter 4, I

call *proto-sentient* or "simple minded"—nor conscious, then, other things being equal, abortion and infanticide are both conventionally morally impermissible with respect to those non-persons. Nevertheless, to carry out an abortion or infanticide in such cases would *not* be a violation of the dignity of the non-person itself, since a non-person does not intrinsically have dignity. Instead, abortion or infanticide in such cases would have a negative impact only on

> (i) the sentient animal life of the non-person, and
> (ii) the moral lives of those members-in-good-standing of The Realm of Ends who rationally support and stand behind the moral convention that constitutes this class of associate members of The Realm, and jointly confer the status of being a secondary subject of dignity and a secondary target of respect upon the erstwhile non-person.

Obviously, the non-person's sentient animal life would be "negatively impacted" by its termination, other things being equal. But there could still be cases in which, even though the abortion or infanticide is *conventionally impermissible*, nevertheless terminating the non-person's life might still constitute *morally permissible euthanasia in a non-conventional sense*. Moreover, the negative impact on the moral lives of those who conventionally confer secondary moral status on the erstwhile non-person, other things being equal, would *not* constitute a violation of their dignity. I particularly stress the *ceteris paribus* qualifier, "other things being," in this connection, however. This is because it is entirely possible, in some contexts, for the abortion or infanticide, actual or threatened, to be *an instrument of coercion directed at those who conventionally confer secondary status on the erstwhile non-person.* For example, if a SWAT team of pro-choicers threatened to abort the first-trimester fetuses of some pro-lifers, in order to force those pro-lifers to vote pro-choice, then that would clearly count as a violation of the pro-lifers' dignity.

It is important to note that associate membership in The Realm of Ends applies every bit as much to living, minimally sentient, or conscious *non*-human non-persons as it does to living, minimally sentient, or conscious *human* non-persons. This in turn makes room, for example, not only for the possibility of an Albert Schweizer-like moral concern for *all* living organisms, and also for any slightly-less-than-Schweizer-like, but still exceptionally broad and inclusive, moral concern over non-sentient human non-person animal organisms like human embryos, or (at best) minimally sentient non-human non-person animal organisms like insects.

The moral convention, according to which secondary dignity and secondary respect is ascribed to human or non-human non-persons, derives ultimately from our respect-based *moral feelings* towards all living, minimally sentient, or conscious non-human creatures in our world, who share with us at least one constitutively necessary feature of real personhood (organismic life), but who are also non-persons because they lack the strong potentiality to become actualized real persons. Indeed, physical nature itself has what I call

proto-dignity, due to its being a constitutive metaphysically necessary condition for the existence of real persons.[164] But neo-personhood, as we have seen, requires the strong-potentiality-of-a-constitutively-necessary-psychological-capacity. A strong potentiality on its own—for example, as possessed by first trimester fetuses after the stage of totipotency—is not enough, and a constitutive psychological capacity on its own—for example, as possessed by sentient non-human non-persons—is also insufficient for neo-personhood.

Associate membership in The Realm of Ends, and its corresponding conventional moral principles, thus both result from coordinated acts of special moral concern and kindness carried out by actualized higher-level or Kantian real human persons, aka persons$_k$, and directed towards any living creatures, especially including minimally sentient animals or non-person conscious animals of any species. And in this way, we can confer associate membership status, secondary dignity, and secondary respect upon, for example, some embryos, fetuses, or infants who are non-persons. This conventional act of conferring a new moral specific character or moral status, in turn, is ultimately justified by reasons that first and foremost determine the morality of our treatment of non-human minded animals.

To summarize what I have been arguing so far. According to The Neo-Person Thesis and its Existential Kantian Ethics-based rationale, neo-personhood is the property of a fetus or infant that determines the difference between the moral permissibility and the moral impermissibility of abortion and/or infanticide. Other things being equal, the possession of neo-personhood or actualized real personhood by a fetus or infant entails the moral impermissibility of abortion or infanticide, and the non-possession of neo-personhood or actualized real personhood by a fetus or infant entails the moral permissibility of abortion or infanticide. And when other things are not equal, there are some special exceptions to this permissibility/impermissibility that derive from the morality of saving and harming others, the morality of one's own death, and the morality of our treatment of non-human animals. *So, holding constant the special exceptions, The Neo-Person Thesis solves the problem of abortion and infanticide.*

It is also significant that The Neo-Person Thesis accurately captures and explains a wide range of commonsense moral intuitions about abortion and infanticide—intuitions that are widely shared by many people who are not extremists (and therefore not partisans) on either side of the philosophical, moral, and political public debate about abortion. These widely shared commonsense, non-extremist moral intuitions include the following:

> (i) other things being equal, fatal abortion is permissible in the first and second trimesters, but impermissible once the fetus is conscious and viable,
> (ii) other things being equal, fatal abortion is permissible in cases of forcible pregnancy due, for example, to rape, even after the stage of fetal consciousness and viability, but if

life-preserving delivery followed by adoption are medically possible, then this is greatly morally preferable or even obligatory,
(iii) other things being equal, fatal abortion is permissible in order to save the life of the mother, and
(iv) other things being equal, fatal abortion and infanticide are both permissible in cases of anencephaly, but infanticide is otherwise impermissible.

The Neo-Person Thesis's conformity to several widely-shared commonsense moral intuitions is not in any way, I hasten to emphasize, *decisive* evidence for the truth of The Thesis. Only conformity to authoritative moral rational intuitions—self-evident and intrinsically compelling moral rational intuitions having an intrinsic connection to moral truth—would be decisive evidence. Still, The Neo-Person Thesis's conformity to widely-shared commonsense moral intuitions, other things being equal, provides some non-trivial rational support for The Thesis. I provisionally conclude, then, that The Neo-Person Thesis is intelligible, defensible, and arguably true, on positive theoretical grounds and on commonsense intuitive grounds alike.

Now I want to argue for The Neo-Person Thesis in a negative way, by considering, comparing, contrasting, and criticizing what I will call *The Standard Approaches* to the morality of abortion and infanticide.

3.5 A CRITIQUE OF THE STANDARD APPROACHES

What I am calling "The Standard Approaches" to the morality of abortion and infanticide fall into four basic distinct kinds:

(i) *The Low Bar of Personhood Approach*,[165] according to which any genetically human creature is a real human person, and therefore has a right-to-life, from the moment of conception or fertilization (that is, during the stage of the embryo or zygote): hence abortion and infanticide are both unrestrictedly impermissible, other things being equal.
(ii) *The Weak Potentiality Approach*,[166] according to which any human creature that has a weak potentiality either for being an actualized real human person or for having a "future like ours"—for example, any normal, healthy embryo which neither divides into twins nor fuses into a chimera during the stage of totipotency—also has exactly the same right-to-life as an actualized real human person: hence abortion and infanticide are both unrestrictedly impermissible for normal, healthy individual human fetuses or infants, other things being equal.
(iii) *The High Bar of Personhood Approach*,[167] according to which only actualized real persons, and, more specifically, only actualized real persons that possess a concept of themselves as continuing desiring, sentient creatures, have a right-to-life: hence abortion and infanticide are both unrestrictedly permissible, other things being equal.

> (iv) *The Right-to-Refuse-Life-Support Approach*,[168] according to which the right-to-life possessed by any creature for whom the mother provides life-support—including human embryos, individual human fetuses, and real human persons—does not include or entail the right of that creature to be provided with whatever it needs in order to go on living, the refusal of which can then be solely determined by the mother's right to control her own body: hence abortion and infanticide are both unrestrictedly permissible, other things being equal.

It is important to note, moreover, that the four standard approaches also all share one fundamental assumption, namely,

> The Right-to-Life Assumption:
>
> If there really is a property whose non-possession or possession by a fetus or infant necessarily determines the moral permissibility or impermissibility of abortion and/or infanticide, then that property always so determines this permissibility or impermissibility via the right-to-life of the fetus or infant.

As I mentioned above, by "the right-to-life" I mean a subject's unalienable moral demand on others to let the subject continue being alive, hence not to be impermissibly actively or passively killed by those others, not a forfeitable right of any sort, and not a strict right-not-to-be-killed. In other words, then, while the four Standard Approaches obviously sharply differ in what they assert about the moral permissibility or impermissibility of abortion and infanticide, nevertheless they all agree in holding that what is of fundamental moral significance is *the right-to-life of the fetus or infant*.

In significant contrast to the four Standard Approaches, The Neo-Person Thesis combines the following four notable features:

> (i) a *middle* bar of personhood approach, based on the notion of a neo-person,
> (ii) a *strong* potentiality, or causal-teleological, approach, based on the notion of a psychological capacity, and
> (iii) a *permission*-to-refuse-life-support approach, using the Existential Kantian Ethics-based idea of the moral impermissibility of harming a person by violating her dignity, other things being equal, together with
> (iv) the notion of associate membership in The Realm of Ends.

By folding a *permission*-to-refuse-life-support approach into its theoretical mix, The Neo-Person Thesis superficially resembles certain critically-refined versions of Thomson's right-to-refuse-life-support argument—for example, those developed by Frances Kamm and David Boonin[169]—both of which appeal to something I will call *the gestational trigger*, that I define as follows:

> *X* is the gestational trigger in the life of a human being *Y* if and only if *X* is some or another developmental stage in the gestational process of *Y* that provides a sufficient condition for ascribing to *Y* a fetal right-to-life (that is, a new unalienable liberty right) which overrides the mother's right-to-refuse-life-support.

Defenders of the refined Right-to-Refuse-Life-Support approach remain openminded about precisely *which* stage will count as the gestational trigger of the fetus's right-to-life, until they have carefully surveyed commonsense moral intuitions, using the classical Rawlsian method of Wide Reflective Equilibrium—which I will discuss explicitly later in this section, and again in section 3.6 below. Now facts about commonsense moral intuitions are either individual facts or social facts. On these refined Right-to-Refuse-Life-Support approaches, then, it is asserted that at some point or another in the gestational process—which remains to be empirically determined by commonsense intuitional individual or social facts—the fetus comes to have a right-to-life that overrides the mother's right-to-refuse-life-support. Hence, other things being equal, abortion is morally permissible prior to the gestational trigger, but morally impermissible after the gestational trigger, holding fixed the now-familiar exceptions for cases of forcible involuntary pregnancy, and self-defense cases in which the mother's life is innocently threatened by the continued existence of the fetus.

By another sharp contrast, for The Neo-Person Thesis, unlike either Thomson's *unrefined* Right-to-Refuse-Life-Support Approach or the *refined* Right-to-Refuse-Life-Support Approach defended by Kamm and Boonin, the rationale for the permission-to-refuse-life-support is *not* based on either the mother's right to control her own body or an empirically-determined gestational trigger of a fetal right-to-life. Instead, according to The Neo-Person Thesis, the rationale for the permission-to-refuse-life-support is based on the specific moral character, or moral status, of the fetus, as guaranteed by the second formulation of the Categorical Imperative (namely, The Formula of Humanity as an End-in-Itself) and the fourth formulation (namely, The Formula of the Realm of Ends), together with The Minded Animalism Theory of personhood and personal identity.

In other words, The Neo-Person Thesis *rejects* The Right-to-Life Assumption shared by all four of The Standard Approaches. According to The Neo-Person Thesis, by sharp contrast, what is of fundamental significance in the morality of abortion and infanticide is not the *rights* of real human persons, but instead the *dignity* of real human persons. Correspondingly, according to The Neo-Person Thesis, what is of fundamental significance are the categorically normative dignity-respecting duties that real human persons have towards other real human persons and towards themselves, duties that they are inherently capable of freely self-legislating. Again, in Existential Kantian Ethics, *human dignity is more fundamental than human rights, and is the moral-metaphysical foundation from which all human rights flow*. Real human persons do indeed have rights—either unalienable or forfeitable rights, and either liberty or claim rights. But rights-based theories

are by their very nature oriented towards how rational human minded animal agents can make moral demands on other agents,[170] whereas for Existential Kantian Ethics the power to make moral demands on other agents flows from dignity and dignity-respecting duties.[171]

Therefore, while I do think that being a subject of dignity and being a target of respect both entail having a right-to-life, nevertheless a right-to-life is *not* what fundamentally matters, morally speaking. And, as per the sub-doctrine of associate membership in The Realm of Ends, a living, minimally sentient, or conscious creature can possess *a merely contingent and extrinsic right-to-life by moral convention* while at the same time still being a non-person and thus lacking dignity. I will return to the critical comparison and contrast between the dignity-based approach and rights-based approaches in section 3.6.

Now I want to spell out explicitly the basic critical arguments against the four Standard Approaches, from the twin standpoints of Existential Kantian Ethics and The Neo-Person Thesis.

(1) Against the Low Bar of Personhood Approach

The Low Bar of Personhood Approach to the morality of abortion and infanticide says that any genetically human creature is a real human person, and therefore has a right-to-life, from the moment of conception or fertilization (namely, the stage of the embryo or zygote), hence abortion and infanticide are both unrestrictedly impermissible, other things being equal. So this Approach sets the moral bar or standard of personhood very low indeed—*every living human animal whatsoever is a human person.* There are at least six good arguments against this approach.

First, during the stage of totipotency, which lasts for approximately 14 to 18 days after conception, twinning and chimeras are still possible. Hence a single embryo can become two (or more) embyros, and two (or more) embyros can become a single embryo. Otherwise put, the identity of embryos is indeterminate. But a real human person has *determinate personal identity conditions*, as spelled out, for example, by The Minded Animalism Theory of personal identity. Therefore, contrary to The Low Bar of Personhood Approach, it is false that any genetically human creature is a human person from the moment of conception or fertilization, and consequently false that any genetically human creature has a right-to-life from the moment of conception or fertilization.

Second, after the totipotency stage but before the emergence of consciousness at 25 to 32 weeks after conception or fertilization, in normal cases, the individual human fetus has none of the constitutively necessary *psychological* capacities of real human personhood. Hence the normal individual human fetus prior to 25 weeks is a non-person. Therefore, contrary to The Low Bar of Personhood Approach, it is false that any normal first trimester fetuses or second trimester human fetuses are human persons, and correspondingly false (leaving aside the possibility of associate membership in The Realm of Ends, which at best extrinsically confers a right-to-life by conventional agreement) that any normal first trimester fetuses or second trimester human fetuses have a right-to-life.

Third, The Low Bar of Personhood Approach prohibits abortion prior to the emergence of consciousness even in cases of forced involuntary pregnancy, for example, rape, and also in cases of accidental involuntary pregnancy, for example, defective birth control devices or techniques. But this entails, by the same reasoning with appropriate changes made for differences in inferential context, that we are morally required to provide life-support for any right-to-life-possessing creature that depends on us for its continued survival, whether or not this involves our being treated as a mere means or as a mere thing—for example, being treated like a "baby machine," in the case of pregnant women—and whether or not we could give our actual or possible rational consent to this. Nevertheless, it is plausibly arguable that we are not morally required to provide life-support for every right-to-life-possessing creature that depends on us for its continued survival. On the contrary, it is plausibly arguable that we are morally required, other things being equal, to provide life support for all and only those right-to-life-possessing creatures for whom we have actually or possibly rationally consented to provide life-support, or to whom we have promised special acts of benevolence or kindness. So in cases in which we forcibly involuntarily become life-support systems for right-to-life possessing creatures, detachment is morally permissible, other things being equal, and thus detachment abortion is morally permissible, other things being equal. Therefore, contrary to The Low Bar of Personhood approach, it is false that abortion is unrestrictedly impermissible, other things being equal.

Fourth, The Low Bar of Personhood Approach prohibits abortion prior to the emergence of consciousness, even in cases in which the mother's life is threatened by the continued existence of the fetus. But this entails, by the same reasoning with appropriate changes made for differences in inferential context, that it is always morally impermissible to defend ourselves against the mortal threats of innocent attackers. On the contrary, however, killing innocent attackers in self-defense is morally permissible, other things being equal, provided that no other way of protecting ourselves is available, and that only minimal lethal force is used. Therefore, again, contrary to The Low Bar of Personhood Approach, it is false that abortion is unrestrictedly impermissible, other things being equal.

Fifth, The Low Bar of Personhood Approach entails that infanticide is unrestrictedly impermissible, other things being equal. But on the contrary, in some cases, for example, the Baby Theresa case, human infants are living but non-sentient non-persons: therefore they lack both dignity and also a dignity-based right-to-life,[172] and can morally permissibly be killed, other things being equal. Moreover, if an infant, whether normal and healthy, or otherwise, is an innocent attacker, then it can be permissibly be killed in self-defense, other things being equal, provided that no other way of protecting ourselves is available, and that only minimal lethal force is used. Therefore, contrary to The Low Bar of Personhood Approach, it is false that infanticide is unrestrictedly impermissible, other things being equal.

Sixth and finally, The Low Bar of Personhood Approach identifies human personhood with being a genetically human creature. But this identification not only confuses being genetically human (in this context, being a stem cell, a sperm, an egg or ovum, or an embryo) with being a real human person (for example, being you or me), it also confuses being genetically human (in this context, being a stem cell, a sperm, an egg or ovum, or an embryo during the roughly 14 to 18 day period of totipotency) with being an individual human animal (that is, in this context, being a human fetus in the period after totipotency and prior to 25 weeks after conception or fertilization). More generally, not all actual or possible rational animals or real persons are genetically human. For one thing, it is arguable that some actual non-human animals are real persons—for example, Great apes, other primates, and perhaps dolphins (see chapter 4 below). And for another thing, it is both analytically conceivable, hence logically and conceptually possible, and also synthetically conceivable, hence really possible, that there are non-human aliens who are real persons—for example, Klaatu in *The Day the Earth Stood Still*,[173] three-headed Zaphod Beeblebrox in *The Hitchhiker's Guide to the Galaxy*,[174] or the mound-people who partially constitute the surface of the planet Quinta in *Fiasco*,[175] to mention only a few of my favorite fictional aliens. Correspondingly, not even all actual human animals are real human persons—for example, anencephalic infants such as Baby Theresa, and human animals in persistent vegetative states such as Karen Ann Quinlan or Terry Schiavo. Therefore, contrary to The Low Bar of Personhood Approach, it is false that being genetically human and being a human person are the same thing.

(2) Against the Weak Potentiality Approach

The Weak Potentiality approach to the morality of abortion and infanticide says that any human creature having a weak potentiality for either being an actualized real human person or having a "future like ours"—for example, any normal, healthy embryo that neither divides into twins nor fuses into a chimera during the stage of totipotency—also has exactly the same right-to-life as an actualized real human person, hence abortion and infanticide are both unrestrictedly impermissible for normal, healthy individual human fetuses or infants, other things being equal. There are at least five good arguments against this approach, the last three of which are essentially the same as the third, fourth, and fifth arguments against The Low Bar of Personhood Approach.

First, by the same weak potentiality principle that gives a right-to-life to the normal healthy embryo that neither divides into twins nor fuses into a chimera during the totipotency stage, so too it follows that the normal healthy sperm and the normal healthy egg or ovum which combine (under favorable biological conditions) to form the embryo that eventually develops into an actualized real human person, also each independently receives a right-to-life. In turn, this has three direct consequences.

(A) All methods of birth control and contraception that involve killing a sperm or ovum are as morally impermissible as arbitrarily killing an actualized real human person.

(B) Even for methods of birth control and contraception that do not involve killing a sperm or ovum, strict celibacy is obligatory whenever one is not having sex for reproductive purposes, since when using these methods, either during non-reproductive sex or afterwards, either healthy spermata or healthy ova are arbitrarily allowed to die, and this is as morally impermissible as arbitrarily allowing a healthy real human person to die.

And (C), it is morally impermissible for most infertile couples to have sex, since either during non-reproductive sex or afterwards, either some healthy spermata or some healthy ova are arbitrarily allowed to die, and this is as morally impermissible as arbitrarily allowing a healthy real human persons to die.

But each of these consequences is morally absurd. Other things being equal, arbitrarily killing a sperm or ovum is obviously not morally equivalent to arbitrarily killing a normal, healthy real human person. Neither a human sperm nor an ovum is even a sentient organism, much less a rational human animal. Moreover, other things being equal, supposing that one were faced with the choice of

> either (i) preventing the arbitrary killing of a normal, healthy sperm or ovum,
> or (ii) preventing the arbitrary killing of a normal, healthy real human person (for example, your next door neighbor),

would there be any even remotely serious question of which creature's arbitrary killing morally deserved to be prevented? Anyone who chose to prevent the arbitrary killing of the sperm or ovum instead of choosing to prevent the arbitrary killing of (for example) his normal, healthy next door neighbor, would be either chillingly evil, plainly insane, or, as neo-Marxists would put it, mind-controlled by a hegemonic ideology.

Correspondingly and for the same basic reasons, other things being equal, arbitrarily allowing a normal, healthy sperm or healthy ovum to die is obviously not morally equivalent to arbitrarily letting a normal, healthy real human person die. Other things being equal, supposing that one were faced with the choice of

> either (i) saving the life of a healthy sperm or ovum,
> or (ii) saving the life of a normal, healthy real human person,

anyone who chose to save the life of the healthy sperm or ovum instead of saving the life of (for example) his normal, healthy next door neighbor, would again be either chillingly evil, plainly insane, or mind-controlled.

Most vividly of all, perhaps, if The Weak Potentiality Approach is correct, then other things being equal, almost no one is ever morally permitted to have non-reproductive sex, on the grounds that arbitrarily allowing a normal, healthy sperm or healthy ovum to die is morally equivalent to arbitrarily letting your normal, healthy next door neighbor die. But that is the moral equivalent of "1=0."

So for all these reasons, the weak potentiality principle is false by a logico-moral reductio argument.[176] Therefore, contrary to The Weak Potentiality Approach, it is false that abortion during the earliest stages of gestation is morally equivalent either to abortion during the later stages of gestation, to infanticide, to arbitrarily killing an actualized real human person, or to arbitrarily letting a real human person die.

Second, The Weak Potentiality Approach fails to distinguish between

> (i) what I have called the *weak potentiality* of a fetus for being an actualized real human person, and
> (ii) what I have called the *strong potentiality* of a fetus for being an actualized real human person.

More specifically, according to Marquis, a "future-like-ours" is the life of a human animal up to its death, *past* a certain designated starting point, beyond which everything is *a non-actual future life for that animal*. Now Marquis argues for the unrestricted moral impermissibility of aborting all normal, healthy human fetuses past the stage of totipotency, on the grounds that it is morally wrong to kill any human animal that has a weak potentiality for having a future life like ours, and in so doing, *deprive it of the counterfactual life that this human animal would have, were it to survive*.

But Marquis's argument is unsound, because he fails to distinguish between

> (i) fetuses that have a future-like-ours but remain non-persons because have not yet acquired what I will call an *actual-life-like-ours*, precisely because they have not yet acquired even the strong potentiality for being an actualized real human person, and
> (ii) fetuses that have a future-like-ours and have *already acquired* an actual-life-ours, that is, fetuses which have already acquired a strong potentiality for being actualized human persons, namely, neo-persons.

Other things being equal, it is not wrong to kill normal, healthy fetuses past totipotentiality but prior to the emergence of their capacity for sentient consciousness, *even despite* their having a future-like-ours, of which they are thereby deprived. Hence Marquis is mistaken about this.

Most explicitly now, for all normal, healthy human animals that have a future-like-ours and have already acquired an actual-life-like-ours—for example, normal healthy fetuses between 25 and 32 weeks after conception or fertilization, normal healthy infants, etc.—then Marquis's approach and The Neo-Person Thesis effectively coincide, at least as far as the ceteris paribus moral impermissibility of arbitrary killing is concerned. But for all normal, healthy human animals that have a future-like-ours, but have *not yet* acquired an actual-life-like-ours—for example, all normal, healthy individual human fetuses prior to 25 weeks after conception or fertilization—then Marquis's future-like-ours argument yields the wrong result. This is because it fallaciously confuses

(i) the moral status of human animals that have a future-like-ours *without* an actual-life-like-ours, with
(ii) the moral status of human animals that have a future-like-ours *together with* an actual-life-like-ours.

Or in other words, Marquis's future-like-ours argument fallaciously converts a *weak potentiality* for actualized real human personhood (that is, mere counterfactual possession of a future-like-ours) into a *strong potentiality* for actualized real human personhood (that is, joint possession of a future-like-ours *and* an actual-life-like-ours). But as against Marquis, The Neo-Person Thesis is claiming that the future life of a fetus, per se, is not what matters morally: only the ongoing life-process of an actual fetus that is already at or past the beginning of its real personal life, is what matters morally. And this point generalizes. In chapter 6, we will see that the ongoing life-process of an actual human animal normatively outranks its future life, not only with regard to the morality of abortion and infanticide, but also with respect to the morality of our own deaths.

Now with application to *every* version of The Weak Potentiality Approach, including of course Marquis's future-like-ours argument, The Neo-Person Thesis is claiming that *only* the strong potentiality of a normal, healthy fetus between 25 and 32 weeks after conception or fertilization, for becoming an actualized real human person—that is, *only* being a creature with an actual-life-like-ours, and hence *only* when this strong potentiality is constituted by the emergence of some of the constitutive psychological capacities of real human persons—will entail that an individual human animal has the same specific moral character or moral status as an actualized real human person, namely its dignity, which in turn entails its dignity-based right-to-life. Therefore, again, contrary to The Weak Potentiality Approach, it is false that abortion during the earliest stages of gestation is morally equivalent either to abortion during the later stages of gestation, or to infanticide, or to arbitrarily killing an actualized real human person.

Third, according to The Weak Potentiality Approach, just like The Low Bar of Personhood Approach, abortion is morally prohibited prior to the emergence of consciousness even in cases of forced involuntary pregnancy, for example, rape, and also in cases of accidental involuntary pregnancy, for example, defective birth control devices or techniques. But this entails, by the same reasoning with appropriate changes made for differences in inferential context, that we are morally required to provide life-support for any right-to-life-possessing creature that depends on us for its continued survival, whether we could give our actual or possible rational consent to this, or not. But it is plausibly arguable that we are not morally required to provide life-support for every right-to-life-possessing creature that depends on us for its continued survival, other things being equal. On the contrary, it is plausibly arguable that we are morally required to provide life support for all and only those right-to-life-possessing creatures for whom we have actually or

possibly rationally consented to provide life-support, other things being equal. So in cases in which we forcibly and involuntarily become life-support systems for right-to-life-possessing creatures, detachment is morally permissible, hence detachment abortion is morally permissible, other things being equal. Therefore it is false, contrary to The Weak Potentiality Approach, that abortion is unrestrictedly impermissible, other things being equal.

Fourth, according to The Weak Potentiality Approach, just like The Low Bar of Personhood Approach, abortion is prohibited prior to the emergence of consciousness even in cases in which the mother's life is threatened by the continued existence of the fetus. But this entails, by the same reasoning with appropriate changes made for differences in inferential context, that it is always morally impermissible to defend ourselves against the mortal threat of innocent attackers, other things being equal. On the contrary, however, killing innocent attackers in self-defense *is* morally permissible, other things being equal, provided that no other way of protecting ourselves is available, and that only minimal lethal force is used. Therefore, again, contrary to The Weak Potentiality Approach, it is false that abortion is unrestrictedly impermissible, other things being equal.

Fifth and finally, The Weak Potentiality Approach, just like The Low Bar of Personhood Approach, entails that infanticide is unrestrictedly impermissible, other things being equal. But in some cases, for example, Baby Theresa, human infants are non-sentient non-persons: therefore, they lack both dignity and also a right-to-life, and can morally permissibly be killed, other things being equal. Moreover, if an infant, whether normal and healthy or otherwise, is an innocent attacker, then it is permissible to kill it in self-defense, other things being equal, provided that no other way of protecting ourselves is available, and that only minimal lethal force is used. Therefore, contrary to The Weak Potentiality Approach, it is false that infanticide is unrestrictedly impermissible, other things being equal.

(3) Against the High Bar of Personhood Approach

The High Bar of Personhood Approach to the morality of abortion and infanticide says that *only* actualized real persons possessing a concept of themselves as continuing desiring, sentient beings, have a right-to-life, hence abortion and infanticide are both unrestrictedly permissible, other things being equal. So this Approach sets the moral bar or standard of personhood very high indeed—*only* self-conscious rational animals are real persons. There are at least three good arguments against this approach.

First, according to The High Bar of Personhood Approach, again, *only* actualized real persons, and, more specifically, *only* actualized real persons that possess a concept of themselves as continuing desiring, sentient beings, have a right-to-life. This entails that having a psychological capacity for self-conscious concept-possession is a necessary condition of being a real human person. Now it is plausibly arguable that a psychological capacity for self-conscious concept-possession also requires a fully online, competent psychological capacity for speaking or understanding a natural language.[177] Hence normal,

healthy third-trimester fetuses after the emergence of consciousness, infants (all of whom are pre-linguistic), and normal healthy toddlers (who possess only minimal linguistic understanding, well short of natural language competence) alike are all *non*-persons by this criterion of personhood.

And in this way, more specifically, arbitrary acts of what I will call *toddlericide* are thereby morally permissible, other things being equal, according to the High Bar of Personhood Approach. But that arbitrary toddlericide should be permissible as a direct consequence of The High Bar of Personhood Approach, is oppositely but equally as morally absurd as the direct consequence of The Weak Potentiality Approach we considered earlier, which says that arbitrarily killing a sperm or ovum is morally equivalent to arbitrarily killing a normal, healthy real human person. Both consequences are the moral equivalent of "1=0." Otherwise put, it is self-evident that normal healthy toddlers are real human persons, and this insight is captured by Part I and Part III of The Extended Four-Part Definition of Real Personhood. Therefore, arbitrary toddlericide is morally impermissible, other things being equal. So by *reductio*, The High Bar of Personhood Approach is false. In turn, contrary to The High Bar of Personhood Approach, it is also false that abortion and infanticide are both unrestrictedly permissible, other things being equal.

Second, The High Bar of Personhood Approach clearly sets the criteria for personhood much too high, and in effect identifies real human personhood with *higher-level* or Kantian rationality and *Kantian real human personhood*, aka personhood_k. But there are, just as clearly, also real human persons that possess only actualized *lower-level* or Humean rationality and therefore only *Frankfurtian real human personhood*, aka personhood_f. This class of real persons especially includes normal healthy toddlers and other normal, healthy older children, all of whom are subjects of dignity and targets of respect, and consequently have a dignity-based right-to-life, yet lack higher-level or Kantian rationality and personhood_k.

Again, from the standpoint of Existential Kantian Ethics, it is self-evident that normal healthy toddlers and other normal healthy older children are actualized real human persons—and so are all normal healthy infants.[178] But let us suppose that you were somehow inclined to doubt that normal healthy *infants* are rational animals in at least the specifically *Humean* or lower-level sense of rationality, so you believe that they are not actualized real human persons, but instead non-persons. Even then, surely, you could not also refuse attributing this moral status to normal, healthy *toddlers* and other normal, healthy *older children*, without falling into moral absurdity. Therefore, again, contrary to The High Bar of Personhood Approach, it is false that abortion and infanticide are both unrestrictedly permissible, other things being equal.

Third, following on from the second critical argument, The High Bar of Personhood Approach fails to recognize either the existence or specific moral character (that is, the moral status)—including the dignity and dignity-based right-to-life—of neo-persons, that

is, normal healthy human fetuses between 25 to 32 weeks after conception or fertilization and afterwards. And it also fails to recognize the moral status of normal healthy infants. In both cases, this is because neither neo-persons nor normal healthy infants possess a concept of themselves as continuing desiring, sentient beings. This entails that normal healthy third-trimester fetuses and normal healthy infants alike can be arbitrarily killed, other things being equal.

But from the standpoint of Existential Kantian Ethics, this is every bit as morally impermissible as the arbitrary killing of normal healthy toddlers, other things being equal, since successful neo-persons and normal healthy infants are all real human persons according to Part I and Part III of The Extended Four-Part Definition of Real Personhood—just like normal healthy toddlers and other normal healthy older children. So by the same token, permitting the arbitrary killing of normal healthy third trimester fetuses and normal healthy infants would also be every bit as morally absurd as permitting the arbitrary killing of normal healthy toddlers (toddlericide) and other normal healthy older children (as it were, *kiddycide*). Therefore, yet again, contrary to The High Bar of Personhood Approach, it is false that abortion and infanticide are both unrestrictedly permissible, other things being equal.

(4) Against the Right-to-Refuse-Life-Support Approach

The *unrefined* Right-to-Refuse-Life-Support Approach to the morality of abortion and infanticide developed by Thomson, says that the right-to-life possessed by any creature for whom the mother provides life-support—including human embryos, individual human fetuses, and rational human minded animals or actualized real human persons (for example, Thomson's famous Violinist)—does not include or entail the right of that creature to be provided with whatever it needs in order to go on living, the refusal of which can then be solely determined by the mother's right to control her own body, Hence, according to this Approach, abortion and infanticide are both unrestrictedly permissible under these conditions, other things being equal. There are at least two good arguments against this approach.

First, if the unrefined Right-to-Refuse-Life-Support Approach is true, then the permissibility of abortion ultimately depends on a forfeitable liberty right of the mother, which the mother can forfeit by, for example, actually or possibly rationally consenting to sex and pregnancy. More precisely, the right to control one's own body, and to protect it from invasion by others, can be forfeited by voluntarily engaging in sex and rationally consenting to an ensuing pregnancy, or by deliberately choosing artificial insemination. In turn, voluntarily engaging in sex and rationally consenting to a follow-up pregnancy, or deliberately choosing artificial insemination, implies a moral commitment to any right-to-life-possessing embryo, fetus, or infant not to kill it "unjustly," which in turn yields a new claim right of the embryo, fetus, or infant against the mother—namely, the right-to-life-support. It then becomes morally impermissible to abort at *any* stage after conception, *even*

during the stage of totipotency, other things being equal. But that seems clearly a mistake, and thus the unrefined Right-to-Refuse-Life-Support Approach is false.

The salient problem here is that for the unrefined Right-to-Refuse-Life-Support Approach, the moral permissibility or impermissibility of abortion is made to depend ultimately on the forfeitable liberty rights of the mother, and on her ability to create or refuse to create new claim rights of the fetus—and not to depend in any salient way on the specific moral character or moral status of the fetus. But according to Existential Kantian Ethics and The Neo-Person Thesis, the morality of abortion is *not fundamentally about either the mother and her liberty rights or her claim-rights-creating choices*, no matter how morally or politically important these rights and choices might otherwise be. Instead, the morality of abortion is *fundamentally about the embryo, fetus, or infant itself and its specific moral character*: namely, its moral status as a non-person, a neo-person, or an actualized real human person. This basic fact is adequately recognized by all Personhood-Based approaches. The issue then becomes whether a low bar, middle bar, or high bar of personhood is the correct way to go. But since, as we have seen, both the low bar and high bar approaches fail, the middle bar approach favored by The Neo-Person Thesis is correspondingly well-supported.

This basic fact is also at least implicitly, or partially, recognized by approaches such as those developed by Kamm and Boonin, grounded on Thomson's original argument in "A Defense of Abortion," that *refine* Thomson's original argument—which, for example, naively identified human beings and persons—by appealing to the (in my terminology, defined above) *gestational trigger* of the fetus's right-to-life. At the same time, Kamm and Boonin also remain open-minded about precisely what will count as the trigger, until they have surveyed commonsense moral intuitions using the Rawlsian method of Wide Reflective Equilibrium.[179] On these approaches, as I mentioned above, it is asserted that at some stage or another in the gestational process, the fetus comes to have a right-to-life—that is, the fetus comes to have an unalienable liberty right constituting a moral demand against others that they let her, namely, the fetus, continue being alive. Once the fetus has this right, then it overrides the mother's right-to-refuse-life-support, which in turn is ultimately determined by the mother's right to control her own body. Then, other things being equal, abortion is morally permissible *prior* to the gestational trigger, but morally impermissible *afterwards*, other things being equal, subject to the now-familiar exceptions for cases of forcible involuntary pregnancy and self-defense cases in which the mother's life is innocently threatened by the continued existence of the fetus. So that explicates the sense in which this is a *refined* Right-to-Refuse-Life-Support Approach.

It seems to me that the refined Right-to-Refuse-Life-Support Approach, all things considered, is the most defensible version of the four Standard Approaches. Indeed, leaving aside some minor differences between Kamm's and Boonin's formulations, the specific moral principles yielded by the refined Right-to-Refuse-Life-Support Approach are at least extensionally equivalent to the first-order substantive ceteris paribus objective moral

principles yielded by The Neo-Person Thesis. My two basic objections to the refined Thomson-style Approach are, instead, meta-ethical.

The first meta-ethical objection stems from a point I argued-for earlier in this section, namely that rights are normatively derivative from dignity and dignity-respecting duties. So if I am correct, then rights-based approaches all *presuppose* The Neo-Person Thesis and Existential Kantian Ethics, or at least they all presuppose some or another reasonable facsimile of The Thesis and Existential Kantian Ethics, that also rely fundamentally on dignity and dignity-respecting duties. Hence in this regard, the refined Right-to-Refuse-Life-Support approach is parasitic on The Thesis and Existential Kantian Ethics.

The second meta-ethical objection is that the refined Right-to-Refuse-Life-Support approach is open to a serious skeptical challenge, which in a nutshell says the following:

> Since the gestational trigger is empirically determined by either individual facts or social facts about commonsense moral intuitions, even when these commonsense moral intuitions have been fully refined and mutually reconciled via the Rawlsian method of Wide Reflective Equilibrium, the fetal right-to-life might be *nothing but a moral convention that is strictly relativized to the relevant individuals or social communities, so "anything goes."*

Or in other words, the refined Right-to-Refuse-Life-Support approach is open to *moral relativism*, which I have already criticized in section 1.2. This worry, in turn, is closely related to a more general criticism of appeals to commonsense moral intuition and the Rawlsian method of Wide Reflective Equilibrium that I will spell out under the heading of Objection 3 in section 3.6 below.

Second, if the unrefined Right-to-Refuse-Life-Support Approach is true, then the following scenario is possible. You rationally consent to a morally permissible detachment abortion during the second trimester, but the fetus survives and is looked after by someone else until it becomes either a neo-person (whether successful or doomed) or else a normal healthy infant or toddler, and thus an actualized real person. Then, one day, you find that very child drowning in a shallow plastic swimming pool on your doorstep. You can then permissibly *refuse* to save its life, merely because you do not want a wet child soiling your nice new front hallway carpet, or for any other self-interested reason.

Thomson says of such cases that they are "morally indecent" (that is, morally bad and scandalous), but *not* "unjust" (that is, morally impermissible).[180] That too seems clearly mistaken, and again it follows that the unrefined Right-to-Refuse-Life-Support Approach is false. As the example I described makes obvious, refusing to save a drowning child in a plastic swimming pool on your doorstep when

> (i) you are the only one who can save it,
> (ii) the act of saving the child costs you nothing of moral significance, even though you do indeed sacrifice something of non-negligible moral value (that is, ruining your nice new front hallway carpet), and

> (iii) you are not morally required to repeat this small act of sacrifice to the point at which it undermines your obligatory life-project, other things being equal, of developing your abilities and perfecting yourself,

is morally equivalent to walking past *a drowning child in a shallow pond* under the same contextual conditions.[181] The further fact that you have previously permissibly detachment-aborted the fetus that became this child is either morally irrelevant to this clear impermissibility, or else makes your act morally even worse, by way of added personal moral responsibility. In both cases, then, the unrefined Right-to-Refuse-Life-Support Approach yields the wrong result.

The problem here, similarly to the problem that was diagnostically identified at the end of the first critical argument against the unrefined Right-to-Refuse-Life-Support Approach, is that the moral permissibility or impermissibility of infanticide is made to depend ultimately on the forfeitable liberty rights of the mother and on her ability to create or refuse to create a new claim right of the infant (namely, the right-to-life-support), and *not* to depend in any salient way on the moral status of the infant itself. But according to Existential Kantian Ethics, just like the morality of abortion, the morality of infanticide is *not fundamentally about the mother of the infant and her liberty rights or her claim-rights-creating choices*, no matter how morally or politically important those rights and choices might otherwise be. The morality of infanticide is *fundamentally about the infant itself and its moral status*.

Again, this basic fact is adequately recognized by Personhood-Based approaches; yet again, the issue then becomes whether a low bar, middle bar, or high bar of personhood is the correct way to go; and *yet* yet again, since both the low bar and high bar approaches fail, the middle bar approach favored by The Neo-Person Thesis is correspondingly well-supported.

3.6 THREE OBJECTIONS AND THREE REPLIES

Now I turn to three possible critical objections against The Neo-Person Thesis, and reply to each of them in turn.

Objection 1: Every Substantive Moral Appeal to Potentiality is Rationally Inadmissible.

The first objection to The Neo-Person Thesis says, in effect, that every substantive moral appeal to potentiality is rationally *bogus*, and that The Neo-Person Thesis substantively appeals to potentiality, therefore The Neo-Person Thesis is rationally inadmissible. The justification for this derives from two basic arguments, one developed by Peter Singer,[182] and one developed by Michael Tooley.[183] The argument developed by Singer, which I will call *Singer's Reductio*, runs as follows:

(1) Any substantive moral appeal to potentiality must entail the following principle: If *X* has the potentiality to become a creature that has a right-to-life, then *X* has a right-to-life. Call this *The Potentiality Principle*.
(2) If The Potentiality Principle is correct, then since a healthy, normal sperm and a healthy, normal ovum each have a potentiality to become a creature that has a right-to-life, it follows that a sperm and an ovum each have a right-to-life.
(3) Other things being equal, either arbitrarily killing a creature that has a right-to-life or arbitrarily letting it die is morally impermissible.
(4) Therefore, other things being equal, either arbitrarily killing a sperm or an ovum or arbitrarily letting it die is morally impermissible.
(5) Since virtually all methods of birth control or contraception, save strict celibacy, involve either the arbitrary killing or arbitrary letting-die of a sperm or an ovum, then it follows that, other things being equal, virtually all methods of birth control or contraception, save strict celibacy, are morally impermissible.
(6) But that is absurd and the moral equivalent of "1=0."
(7) So The Potentiality Principle is false.
(8) Therefore any substantive moral appeal to potentiality is rationally inadmissible.

Correspondingly the argument developed by Tooley, which I will call *Tooley's Counterexample*, runs as follows:

(1) Any substantive moral appeal to potentiality must entail the following principle: If *X* has the potentiality to become a creature that has a right-to-life, then *X* has a right-to-life. Call this The Potentiality Principle.
(2) But there are counterexamples to The Potentiality Principle. In order to show this, we must first accept *The Moral Symmetry Principle*:

Let *C* be a causal process that normally leads to outcome *E*. Let *A* be an action that initiates process *C*, and let *B* be an action involving a minimal expenditure of energy that stops process *C* before outcome *E* occurs. Assume further that actions *A* and *B* do not have any other consequences, and that *E* is the only morally significant outcome of process *C*. Then there is no difference between intentionally performing action *B* and intentionally refraining from performing action *A*, assuming identical motivation in both cases.

(3) Granting The Moral Symmetry Principle, here is a counterexample to The Potentiality Principle:

If a serum were developed that turns kittens into persons, then we would not think it wrong to refuse to give kittens the serum. But since, by The Moral Symmetry Principle, there is no moral difference between refusing to give the kittens the serum and intervening to stop their process of development into persons once we have given it to them, then it is permissible to kill the injected kittens. Hence the kittens do not have a right-to-life. So even though the kittens have the potential to become a creature that has a right-to-life, they do not thereby have a right-to-life.

(4) So The Potentiality Principle is false.
(5) Therefore any substantive moral appeal to potentiality is rationally inadmissible.

Two Replies to Objection 1

First, Both *Singer's Reductio* and *Tooley's Counterexample* share premise (1), which says that any substantive moral appeal to potentiality must entail The Potentiality Principle. But this premise is false. The Neo-Person Thesis carefully distinguishes between *weak potentiality* and *strong potentiality*, and correspondingly makes a substantive moral appeal only to *the strong-potentiality-of-a-constitutively-necessary-psychological-capacity for being a real human person*. By contrast, The Potentiality Principle relies on *weak* potentiality, not on *strong* potentiality. Hence the most that could follow from either *Singer's Reductio* or *Tooley's Counterexample* is that all substantive moral appeals to *weak potentiality* are rationally inadmissible. Indeed, as we have seen, one of my critical arguments against The Weak Potentiality Approach itself deploys a version of *Singer's Reductio*. So The Neo-Person Thesis is unaffected.

Second, *Tooley's Counterexample* is unsound because The Moral Symmetry Principle itself has counterexamples, and is therefore false. The Moral Symmetry Principle says, in effect,

(i) that if it is morally permissible to refuse to start a process, then it is morally permissible to intervene at any point in the process and stop it before it ends, and also
(ii) that if it is morally obligatory to start a process, then it is morally obligatory to let the process run through to the end.

Now let us consider Tooley's kittens again. Being rationally charitable to Tooley's line of reasoning, we can assume that

(i) the kittens start with a psychological capacity for consciousness,
(ii) it is not seriously morally wrong to kill untransformed kittens, other things being equal (but see also chapter 4 below, for an unrestricted moral prohibition against torturing minded animals), and
(iii) it is not wrong to refuse to give the kittens the serum, other things being equal.

So if we kill the kittens before we give them the serum, or if we kill them as soon as we give them the serum but nothing has happened to them yet, then those are both morally permissible, other things being equal.

But suppose now that instead of killing the kittens, either before we give them the serum or as soon as we give them the serum but nothing has happened to them yet, we let the serum begin to transform them. Then the first salient thing that happens, by the hypothesis of Tooley's thought-experiment, is that the transformed kittens begin to manifest *a real person's psychological capacity for consciousness*, and not *merely a*

kitten's psychological capacity for consciousness. Now in view of the fact that feline bodies are sharply different from human bodies, we can plausibly suppose that the difference between the untransformed kittens' subjective experience and the transformed kittens'subjective experience is going to be almost as sharp as the radical difference, or "mental-mental gap," between a *bat's* subjective experience and *our* subjective experience—which, in turn, is a basic premise in Thomas Nagel's famous argument for the non-reducibility of mentalistic concepts to physicalistic concepts, aka the "mental-physical gap" argument.[184] So, given the emergent fact of the new and sharply different psychological capacity manifested by the transformed kittens, together with the further strongly potential fact that the transformed kittens will become actualized persons in the natural course of their later neurobiological development, it follows that the transformed kittens are non-human neo-persons. By The Neo-Person Thesis, it then follows that the transformed kittens are subjects of dignity and targets of respect. Therefore, they also possess a dignity-based right-to-life, and it is now morally impermissible to kill them arbitrarily, other things being equal. Therefore The Moral Symmetry Principle is false, and *Tooley's Counterexample* is unsound.

This critical result does not, of course, vindicate The Potentiality Principle, which as we have seen is false on independent grounds, along the lines of *Singer's Reductio*. But the unsoundness of *Tooley's Counterexample* does indeed vindicate The Neo-Person Thesis's substantive moral appeal to the strong-potentiality-of-a-constitutively-necessary-psychological-capacity for being a real human person.

Objection 2: The Concept PERSON is an Ambiguous, Incomplete, and Vague Concept

This second objection trades on a deep and widely-held skepticism in recent and contemporary professional academic philosophy, about the nature of concepts and also about conceptual analysis as a philosophical method. Moreover, this skepticism has also been famously directly applied to the concept PERSON by Daniel Dennett and Mary Anne Warren.[185] Their well-known accounts each entail that the concept PERSON is

> (i) *ambiguous* in the sense that there are several different and outright inconsistent or at least incommensurable conceptions of personhood,
> (ii) *incomplete* in the sense that even if some *necessary* conditions of personhood can be found—a thesis explicitly defended by both Dennett and Warren—nevertheless no list of such conditions can be found such that it constitutes a universally necessary and sufficient condition for personhood, and also
> (iii) *vague* in the sense that there are actual or possible cases of creatures, relative to any proposed list of conditions for personhood, that are neither strictly speaking persons nor strictly speaking non-persons.

Now since, according to the Dennett-Warren line of reasoning, the concept PERSON is an ambiguous, incomplete, and vague concept, and since The Neo-Person Thesis centrally and explicitly deploys the concept REAL PERSON and its proposed explicit real

definition, therefore The Neo-Person Thesis must be an ambiguous, incomplete, vague, unintelligible, and indefensible doctrine.

Two Replies to Objection 2

First, it seems obvious that, merely because *some* accounts of the concept PERSON entail that it is an ambiguous, incomplete, and vague concept, it does *not* thereby follow that *all* accounts of the concept PERSON entail this. Moreover, even if it is true that the *determinable* concept PERSON is in some respects ambiguous, incomplete, and vague, it does not follow that all of its *determinates* are ambiguous, incomplete, and vague. Indeed, The Minded Animalism Theory of personhood, if correct, explicitly entails that the determinate concept REAL PERSON is neither ambiguous, incomplete, nor vague. So it is simply not true that every account of the concept PERSON entails a skeptical conception of that concept. Correspondingly, The Neo-Person Thesis is cogent to the extent that The Minded Animalism Theory is cogent.

Second, in view of The Minded Animalism Theory of personhood, there are good reasons for holding that the determinate concept REAL PERSON is *neither* ambiguous, *nor* incomplete, *nor* vague, once the following distinctions have been made:

(i) personhood vs. personal identity,
(ii) personhood per se vs. real personhood,
(iii) Frankfurtian real persons vs. Kantian real persons, and
(iv) non-persons vs. neo-persons vs. actualized real persons.

Indeed, as I have argued earlier in this chapter, these distinctions make it possible to provide an explicit definition of real personhood, via a set of individually necessary, individually insufficient, and jointly sufficient conditions for being a real person.

In this connection it is extremely important to note that *the semantic structure of a concept* is distinct from *the epistemology of concept-application in judgments*. This is a basic lesson that emerges in a thoroughgoing critique of Quine's attack on the analytic-synthetic distinction.[186] And it is also the basic rationale lying behind my distinction between the semantics of objective moral principles and the epistemology of moral judgments in section 2.1 above. Hence, even if it is extremely difficult, in some cases, to know with certainty whether a certain concept applies to a certain case, it simply does *not* follow that there is any ambiguity, incompleteness, or vagueness in the semantic structure of the concept (or any corresponding objective moral principle based on that concept) itself. Hence it is perfectly legitimate for The Neo-Person Thesis to rely on the explicit definition of the concept of real personhood provided by The Minded Animalism Theory.

Objection 3: The Neo-Person Thesis Fails to Proceed Fundamentally Via the Right-to-Life

The third objection, in effect, deploys The Right-to-Life Assumption as a dialectical weapon, and says that

(i) any acceptable approach to the morality of abortion and infanticide *must* proceed fundamentally via the right-to-life,
but (ii) The Neo-Person Thesis does *not* proceed via the right-to-life,
therefore (iii) The Neo-Person Thesis is unacceptable.

Two Replies to Objection 3

First, and almost self-evidently, the very fact that The Right-to-Life Assumption is an *assumption*, and not a *proven assertion*, makes it possible for The Neo-Person Thesis to be defended by simply rejecting that assumption. This rejection, in turn, is justified by the fact that it simply has *not* been proven by the defenders of The Standard Approaches that every acceptable approach to the morality of abortion and infanticide *must* proceed fundamentally via the right-to-life of the fetus. On the contrary, and as we have already seen in detail in section 3.5, what on the contrary very clearly appears to be the case, is that every approach to the morality of abortion and infanticide which adopts The Rights Assumption—or at least, each of the four Standard Approaches—is open to decisive objections against it.

This, in turn, strongly suggests that The Rights Assumption is false. There is something morally more fundamental than the right-to-life, namely the dignity of real human persons and the dignity-respecting duties of rational human moral agents. I will re-emphasize that I am *not* saying that real human persons do *not* have a right-to-life—on the contrary, I am saying that they *do* have a right-to-life. What I am saying is simply that the dignity of real human persons and the dignity-respecting duties of rational human moral agents are morally more fundamental than the right-to-life. Human rights flow from human dignity, and not the converse.

Second, following on from that point, even when we focus on the most defensible version of The Standard Approaches—namely, the refined Right-to-Refuse-Life-Support Approach—there remains at least one basic objection to it. As I mentioned above, this basic objection is a meta-ethical skeptical challenge which starts from the fact that according to the refined Right-to-Refuse-Life-Support approach, the (in my terminology, defined above) *gestational trigger* is empirically determined by individual facts or social facts about commonsense moral intuitions, even when these are fully refined and mutually reconciled by the Rawlsian method of Wide Reflective Equilibrium. Hence the fetal right-to-life might be nothing but a mere moral convention that is strictly relativized to the relevant individuals or social communities.

To the extent that this entails moral relativism, whether at the level of individuals (individual relativism) or social communities (cultural relativism), then it follows that the refined Right-to-Refuse-Life-Support Approach is open to the critical arguments against moral relativism that I presented in section 1.2. So I will not repeat those arguments here.

Leaving aside those worries about moral relativism, the crucial meta-ethical difference between The Neo-Person Thesis and the unrefined Right-to-Refuse-Life-Support approach

has to do with the latter's methodological commitment to commonsense moral intuitions and the Rawlsian method of Wide Reflective Equilibrium. The Neo-Person Thesis is grounded in a robust essentialist moral metaphysics. So its conclusions are, if true, then necessarily true; and the moral distinctions that it draws, if they are correct distinctions, accurately track and necessarily flow from inherent structures in The Web of Mortality. But the refined Right-to-Refuse-Life-Support approach is based *only* on commonsense moral intuitions and reflective equilibrium, and *nothing else*. I specially emphasize the "only" and the "nothing else." This is because The Neo-Person Thesis *also* takes into account commonsense moral intuitions and reflective equilibrium as evidence. That in turn accounts for the partial overlap between The Neo-Person Thesis and the refined Right-to-Refuse-Life-Support approach, and the extensional equivalence of their first-order substantive ceteris paribus moral principles. Nevertheless, for The Neo-Person Thesis, commonsense moral intuitions and reflective equilibrium are treated only as *prima facie* data and *prima facie* evidence for substantive conceptual and metaphysical distinctions, that must also be independently justified. Hence it is entirely possible that commonsense moral intuitions and/or what results from reflective equilibrium will ultimately be rejected on substantive conceptual or metaphysical grounds.

Otherwise put, without this substantive conceptual and metaphysical grounding, it always remains possible that the commonsense intuitive data and evidence, even when fully refined and harmonized by reflective equilibrium, will track merely contingent facts about idiosyncratic, illusory, or ideologically "mind-controlled" beliefs of individuals or social communities, and not inherent structures in The Web of Mortality.

This contingent openness to individual or social idiosyncrasy, illusion, and ideological mind-control, in turn, also afflicts The Rights Assumption more generally, since rights are generally held to be conventional facts. Now it is of course true that classical theories of rights generally attempt a grounding and a stronger constraining in terms of *natural rights*, often with a theological foundation. But on the one hand, a commitment to the existence of natural rights is *not* a commitment of either the unrefined or refined Right-to-Refuse-Life-Support approaches to the morality of abortion and infanticide. And on the other hand, all classical natural rights theories fall under either The Weak Potentiality approach or The Low Bar of Personhood approach,[187] both of which, as we have already seen, are open to decisive objections.

3.7 CONCLUSION

In this chapter I have argued that The Neo-Person Thesis, aka The Thesis, is defensible on positive theoretical, commonsense intuitive, and critical-dialectical grounds alike, and also that, on behalf of Existential Kantian Ethics, The Thesis adequately solves the basic problems facing the four Standard Approaches to the morality of abortion and infanticide.

The Thesis provides robust essentialist conceptual and metaphysical foundations for the morality of abortion and infanticide. These deep foundations, in turn, guarantee that The Thesis is adequately buffered against serious skeptical worries that can afflict even the most defensible rights-based approaches. Moreover, it should be noted explicitly that in a contemporary American sociopolitical sense, The Thesis is in some crucial respects "pro-life" and in some other crucial respects "pro-choice," although at the same time it is neither *one-sidedly* pro-life nor *one-sidedly* pro-choice. Thereby it effectively avoids the ideologically-driven, false PRO-LIFE vs. PRO-CHOICE dichotomy. Of course, it would be spurned by cognitively-blinkered ideologues in both camps—if they ever looked up from their highly-filtered FaceBook pages, highly-hashtagged Twitter feeds, texting, tweeting, etc., etc, and paid any attention whatsoever to it, that is—but that's just a regrettable fact of contemporary life in the USA.

All things considered, then, The Neo-Person Thesis, based on Existential Kantian Ethics, constitutes a conceptually and metaphysically well-grounded, and true, but also socially and politically subtle, if not popular, approach to the morality of abortion and infanticide.

Chapter 4

WHAT IS IT LIKE TO BE A BAT IN PAIN? THE MORALITY OF OUR TREATMENT OF NON-HUMAN ANIMALS

With regard to the animate but nonrational part of creation, violent and cruel treatment of animals is far more intimately opposed to a human being's duty to himself [than a propensity to the destruction of what is beautiful in inanimate nature], and he has a duty to refrain from this; for it dulls sympathy in the human being for their pain and so weakens and gradually uproots a natural predisposition that is very serviceable to morality in one's relations with other people. The human being is authorized to kill animals quickly (without pain) and to put them to work that does not strain them beyond their capacities (such work as himself must submit to). But agonizing physical experiments for the sake of mere speculation, when the end could be a achieved without these, are to be abhorred. –Even gratitude for the long service of an old horse or dog (just as if they were members of the household) belongs *indirectly* to a human being's duty *with regard to* these animals; considered as a direct duty, however, it is always only a duty of the human being *to* himself. (*MM* 6: 443)

We describe bat sonar as a form of three-dimensional forward perception; we believe that bats feel some versions of pain, fear, hunger, and lust, and that they have other, more familiar types of perception besides sonar. But we believe that these experiences also have in each case a specific subjective character, which it is beyond our ability to conceive.[188]

The day *may* come when the rest of the animal creation may acquire those rights which never could have been witholden from them but by the hand of tyranny. The French have already discovered that the blackness of the skin is no reason why a human being should be abandoned without redress to the caprice of a tormentor. It may one day come to be recognized that the number of legs, the villosity of the skin, or the termination of the *os sacrum* are reasons equally insufficient for abandoning a sensitive being to the same fate. What else is it that should trace the insuperable line? Is it the faculty of reason, or perhaps the faculty of discourse? But a full-grown horse or dog is beyond comparison a more rational, as well as a more conversable animal, than an infant of a day or a week or even a month old. But suppose it were otherwise, what would it avail? The question is not, Can they *reason?* nor Can they *talk?* but, *Can they suffer?*[189]

> If a being suffers, there can be no moral justification for refusing to take that suffering into account… If a being is not capable of suffering, …, there is nothing to be taken into account.[190]

4.1 INTRODUCTION

How ought we to treat non-human minded animals? According to Existential Kantian Ethics, the correct answer to this question flows ultimately from *the nature of the subjective experience of pain in different types of minded animals*—including minded animals not only as radically strange as bats or octopuses, but also as all-too-familiar as ourselves, namely, real human persons, alike.

More specifically, however, the problem I am focusing on here is whether non-human minded animals, like bats or cats, subjectively experience the same *kind* of pain as real human persons, or not, and what the moral implications of the answer to that question are for the morality of our treatment of non-human animals, whether sentient and fully minded (like bats or cats) or proto-sentient and "simple minded" (like cephalopods, fish, insects, or reptiles). Let us call this *The What-Is-It-Like-To-Be-A-Bat-In-Pain?Problem*. In what follows in this chapter, against the dual backdrop of Existential Kantian Ethics and The Minded Animalism Theory of personhood and personal identity,[191] I grapple with The What-Is-It-Like-To-Be-A-Bat-In-Pain? Problem, and propose a comprehensive solution to it. This comprehensive solution is given by an explication and defense of the following six moral theses or first-order substantive ceteris paribus objective moral principles:

> (1) *The Non-Speciesist Real Person Thesis*. Some but not all human beings are real persons, and some but not all non-human beings are real persons.
> (2) *The Bodily Pain vs. Suffering Thesis*. All minded animals, whether human or non-human, are capable of experiencing bodily pain, or in other words, all minded animals are capable of what I call *bodily nociperception* (see section 4.1 below). But all and only real persons, including all rational human minded animals and also some rational but non-human minded animals, are capable of subjectively experiencing specifically *emotional* pain—that is, all and only real persons are capable of *suffering*.
> (3) *The Pain-and-Suffering Principle*. The suffering of any real person is always a primary target of serious moral concern for every higher-level or Kantian real human person. Moreover, as higher-level or Kantian real human persons, we always have good reason to fear our own future suffering, other things being equal. Most importantly, however, other things being equal, every higher-level or Kantian real human person is obligated never to treat any real person, whether human or non-human, in such a way as to cause them to suffer by violating their dignity, namely, to *degrade* them. Finally, the experience of pain by minded animals of any species is also always a target of serious moral concern for higher-level or Kantian real human persons.

(4) *The Moral Comparison Thesis*. There is compellingly good reason to believe that the suffering of any human or non-human minded animal that *is* a real person, whether via bodily nociperception or without bodily nociperception, is substantially more morally significant than the bodily nociperception of any human or non-human minded animal that is *not* a real person, assuming roughly comparable degrees of experienced intensity.
(5) *The Bodily-Pain-without-Torture-or-Cruelty Principle.* Other things being equal, it is morally permissible for higher-level or Kantian real persons to treat either human or non-human minded animals that are non-persons in such a way that it foreseeably causes some state of bodily nociperception in them, although it is morally impermissible to *torture* them, or treat them with *cruelty*.
(6) *The Associate Membership Thesis*. Higher-level or Kantian real human persons can create moral conventions for treating selected groups of human or non-human, minded or non-minded non-persons temporarily or permanently *as if* they were real human persons falling under the protection of the Categorical Imperative, provided that those creatures are, at least, individual living organisms; and as a consequence, those human or non-human, minded or non-minded non-persons thereby gain an "associate membership in The Realm of Ends," whereby they are secondary subjects of dignity and secondary targets of respect, and thus receive a temporary or permanent right-to-life.

The conjunction of these six theses or principles is what I will call *The Concern For All Minded Animals Theory* of the morality of our treatment of non-human minded animals, since it entails

not only that (i) *some* non-human minded animals, as real persons, inherently are subjects of dignity and targets of respect,
but also that (ii) *all* non-human minded animals inherently are experiencers of moral value and targets of moral concern,
and also that (iii) *all* non-human minded animals, whether sentient and fully minded (like bats or cats) or proto-sentient and "simple minded" (like cephalopods, fish, insects, or reptiles), or even non-human, non-minded individual living organisms (like early-stage bat-fetuses or cat-fetuses), extrinsically considered, are possible targets of moral concern in relation to possible associate membership in The Realm of Ends.

It should be particularly noted that The Concern For All Minded Animals Theory cuts sharply across the familiar division—often assumed to be exhaustive—between the anthropocentric (aka "speciesist")[192] and *anti*-anthropocentric (aka "anti-speciesist")[193] normative ethical positions. Anthropocentrism or speciesism says that biological species membership is the sole or at least the primary determinant of moral distinctions between creatures; and anti-anthropocentrism or anti-speciesism says that species membership is wholly irrelevant to moral distinctions between creatures. In turn, the famous or notorious *Animal Liberation* view defended in different ways by Tom Regan and Peter Singer,[194] which says that all sentient non-human animals deserve equality of moral consideration

and/or treatment with persons,[195] is a sub-species of anti-speciesism. Nevertheless, it is fully consistent to reject speciesism while still holding that species membership *partially* determines moral distinctions between creatures, or at the very least that species membership is *significantly relevant* to moral distinctions between creatures. So The Concern For All Minded Animals Theory is *neither* anthropocentrist *nor* anti-anthropocentrist, and in turn, *neither* speciesist *nor* anti-speciesist. More precisely, according to The Concern For All Minded Animals Theory,

> (i) real persons are subjects of dignity and targets of respect, no matter what species they belong to,
> (ii) some but not all human animals are subjects of dignity and targets of respect,
> (iii) some but not all non-human minded animals are subjects of dignity and targets of respect,
> (iv) all minded animals of any species are experiencers of moral value and targets of our moral concern, and
> (v) other things being equal, the morally permissible treatment of any minded animal of any species is determined by the kind of pain it can experience, which in turn is partially determined by the neurobiology of the species that it belongs to.

In this way, according to The Concern For All Minded Animals Theory, the morality of our treatment of non-human minded animals *neither* strictly tracks differences between biological species, *nor* does it favor humans, *nor* does it wholly ignore species differences.

It is point (v), perhaps, that will be most surprising, since this is the one that directly expresses *a non-speciesist moral appeal to species differences*, via the concept of the experience of pain in minded animals. This appeal is captured in theses or principles (1) to (6). In sections 4.2 to 4.4, I will argue for each of these theses or principles in turn. Then in section 4.5, I will argue for The Associate Membership Thesis, which again directly expresses a non-speciesist moral appeal to species differences, but in a sharply different way, in that it captures the special moral concern or *kindness* that higher-level or Kantian real human persons can actively express for any kind of minded animal whatsoever, whether belonging to its own species or any other species, or indeed for any kind of any living organism.

4.2 REAL PERSONS AND DIFFERENT SPECIES

Here, again, is the extended, four-part definition of real personhood that I worked out in section 3.3 above:

The Extended, Four-Part Definition of Real Personhood

Part I. X is a real Frankfurtian person (person$_f$) if and only if X is an S-type animal and X has fully online psychological capacities for:

(1) *essentially embodied consciousness* or essentially embodied subjective experience,
(2) *intentionality* or directedness to objects, locations, events (including actions), other minded animals, or oneself, including cognition (that is, sense perception, memory, imagination, and conceptualization), desire-based emotions, and effective first-order desires,
(3) *lower-level or Humean rationality*, that is, logical reasoning (including judgment and belief) and instrumental decision-making,
(4) self-directed or other-directed *evaluative emotions* (for example, love, hate, fear, shame, guilt, pride, etc.),
(5) *minimal linguistic understanding*, that is, either inner or overt expression and communication in any simple or complex sign system or natural language, including ASL, etc., and
(6) second-order volitions.

Part II. X is a real *Kantian person* (person$_k$) if and only if X is a real person$_f$ and also has fully online psychological capacities for:

(7) *higher-level or Kantian rationality*, that is, categorically normative logical rationality and practical rationality, the latter of which also entails a fully online capacity for autonomy (self-legislation) and wholeheartedness, hence a fully online capacity for principled authenticity.

Part III. X is a real person if and only if X is either a real person$_f$ or a real person$_k$, otherwise X is a *non-person*.

Part IV. If X is an actualized real person, then the neo-person of X is also a real person, where the neo-person of X is a given individual S-type animal A that manifests the psychological capacity for consciousness and the following counterfactual is also true of A:

If A *were* to continue the natural course of its neurobiological and psychological development, then *A would* become X.

Given some familiar facts about human animals, it follows from The Extended, Four-Part Definition of Real Personhood that not all human beings are real persons. For example, normal, healthy fetuses past the stage of totipotency but prior to the emergence of full sentience (that is, prior to approximately 25 weeks in the gestation period), anencephalic fetuses and infants, and human beings in persistent vegetative states, all lack a capacity for

consciousness, and therefore are non-persons under both Part I and Part IV of the extended, four-part definition of real personhood.

At the same time, however, in view of strong evidence from cognitive ethology,[196] then at least some non-human animals—and in particular, Great apes, other primates, and perhaps dolphins—are in fact real persons under Part I and Part III of the extended, four-part definition of real personhood. More precisely, at least some non-human animals, including Great apes, other primates, and perhaps dolphins, are real persons precisely because they are *Frankfurtian* persons, aka persons$_f$. There is good evidence that these non-human animals have online psychological capacities for consciousness, intentionality, lower-level or Humean rationality, self-directed or other directed evaluative emotions, minimal linguistic understanding, and second-order volitions. If so, then they are intentional agents who are thereby capable of what I call *free volition*,[197] even if they are not strictly speaking capable of what I call *free agency*—that is, the conjunction of free will and practical agency[198]—which includes the morally high-powered innately specified capacity for achieving principled authenticity, at least partially or to some degree. So, as far as the available evidence indicates, there are no non-human minded animals whose fully online capacities put them within reach of principled authenticity, even if at least some of them are rational minded animals or real persons possessing absolute, non-denumerably infinite, intrinsic, objective moral value, namely, dignity.

In this way, however, real persons who are also non-human minded animals are *primary subjects of dignity and primary targets of respect*, because they fall directly under the Categorical Imperative, and therefore they must be both considered and treated as such, even though they belong to different species. It is impermissible to treat them either as mere means or as mere things, and/or without their actual or possible rational consent—that is, to treat them without respect—since this would harm them by violating their dignity. To treat a Great ape, other primate, or perhaps a dolphin, either as a mere means or as a mere thing, and/or without its actual or possible rational consent, *would be just like treating a normal, healthy human toddler or other normal, healthy child either as a mere means or as a mere thing, and/or without her actual or possible rational consent*. This is not to say that Great apes, other primates, or dolphins are neurobiologically or psychologically *interchangeable*, or *intersubstitutable*, with normal, healthy toddlers or other normal, healthy human children, but rather just that they do share with normal, healthy toddlers and other normal, healthy human children the same set of constitutively necessary psychological capacities, and the same moral specific character or moral status. We are morally obligated to care morally about them in the same way, and to treat them in the same way, that we do normal, healthy toddlers and other normal, healthy human children.

Put somewhat trivially, but still relevantly and perhaps also vividly, this moral obligation accounts, for example, for the undeniable emotional and moral impact of the classic 1933 thriller *King Kong*.[199] You feel deeply sorry for The Big Ape, deeply

sympathetic with his obvious love for the Fay Wray character Ann Darrow, and morally outraged by what they have done to him. In the context of the movie, it is clear that King Kong is a morally much better real person than the Robert Armstrong character Carl Denham, the ambitious and heartless promoter.

Less trivially now, real persons who are also non-human animals should not be subjected to any medical or scientific experimentation, unless it is precisely the sort of medical or scientific experimentation that is morally permissible for normal healthy toddlers and other normal healthy human children. In other words, other things being equal, we morally must not torture or vivisect normal healthy toddlers or other normal healthy human children in the name of Medicine or Science: therefore, other things being equal, neither should we torture or vivisect Great apes, other primates, or perhaps dolphins in the name of Medicine or Science. Furthermore, other things being equal, rational minded animals or real persons who are also non-human animals should not be kept in zoos, or in any other sort of captivity, unless it can be clearly shown that this is what they naturally need or rationally want.

Like normal healthy toddlers and other normal healthy human children, who both naturally need and rationally want to be looked after, it might well be that rational minded animals or real persons who are also non-human animals may sometimes also naturally need or rationally want to be looked after. Indeed, real human persons who are also fully higher level or Kantian real persons, or persons$_k$, sometimes naturally need and rationally want to be looked after too: for example, by their loved ones under normal conditions; in hospitals when they are sick; or in managed care apartments, or hospices, etc., when they get old and need constant attention. But normal healthy toddlers and other normal healthy human children neither naturally need nor rationally want to be kept in *zoos* or other sorts of cages. Keeping a normal healthy toddler or other normal healthy human child in a zoo or any other sort of cage is clearly morally impermissible, and would be treating them as mere means or mere things, and/or without their actual or possible rational consent. That would be acting like a Nazi, or like an evil character right out of the fairly scary Brothers Grimm version of *Hansel and Gretel*, or the (to me) heart-stoppingly scary horror film, *The Blair Witch Project*.[200] Correspondingly, then, with appropriate modifications made for change of context, the same goes for real persons who are also non-humans.

As we have just seen, the available evidence strongly indicates that some non-human minded animals are real persons. But assuming that this is true, where does it leave all the *other* non-human minded animals? By Part III of the extended, four-part definition of real personhood, anything that is not a real person is a non-person. But are all non-persons the same, morally speaking? Are all non-persons equivalent to mere things? *No*. This is because all mere things are natural mechanisms, but some non-persons are living, sentient organisms, that is, minded animals. So minded non-human animals that are also non-persons are *not* morally equivalent to mere things. All minded animals—even proto-sentient or "simple minded" non-human animals like cephalopods, fish, insects, reptiles,

and other invertebrates, but especially including all sentient and fully minded non-human animals like bats, bears, birds, cats, cows, dogs, horses, lions, mice, sheep, and wolves—are experiencers or primary subjects of moral value, and also primary targets of our moral concern, even if they are non-persons.

According to Existential Kantian Ethics, moral values are in the world because minded animals are in the world, and all ethical values necessarily depend on moral values as their essence. In other words, according to Existential Kantian Ethics, all minded living organisms must be considered individually, and each of them must be taken fully into account in our moral reasoning—even if they are not thereby morally considered or treated *equally* as members of the universal intersubjective moral community of real persons, The Realm of Ends. This moral concern for all minded animals is determined by the fact that they all share with us at least two constitutively necessary conditions of real personhood, namely organismic life and (proto-) sentience, both of which are necessarily contained within, and thus partially constitutive of, essentially embodied consciousness. In sections 4.3 and 4.5 below, we shall see that (proto-) sentience in a minded animal carries with it the psychological capacity for experiencing pain, and also that this provides a serious target for our moral concern.

But right now we need to get somewhat clearer on the notion of a "minded animal." As I noted earlier, the dictionary meaning of the word 'animal' is "a living organism which feeds on organic matter, usually one with specialized sense organs and nervous system, and able to respond rapidly to stimuli."[201] In biology on the other hand, 'animal' has a more technical meaning, in that animals constitute one of the five kingdoms of living things: Monera (bacteria), Protists, Fungi, Plants, and Animals. The class of animals in this biological sense includes both vertebrates and invertebrates. So my usage of 'animal' in this book, as in *Embodied Minds in Action* and *Deep Freedom and Real Persons* alike, is a precisification of the ordinary language and scientific terms, intended to coincide with its normal use in cognitive ethology. To signal this precisification, I coined the quasi-technical term *minded animal*.

By the notion of a "minded animal," again, I mean any living organism with inherent capacities for

> (i) *consciousness*, that is, a capacity for embodied subjective experience,
> (ii) *intentionality*, that is, a capacity for conscious mental representation and mental directedness to objects, events, processes, facts, acts, other animals, or the subject herself (so in general, a capacity for mental directedness to *intentional targets*), and also for
> (iii) *caring*, a capacity for conscious affect, desiring, and emotion, whether directed to objects, events, processes, facts, acts, other animals, or the subject herself.

Over and above consciousness, intentionality, and caring, in some minded animals, there is also a further inherent capacity for

> (iv) *rationality*, that is, a capacity for self-conscious thinking according to principles and with responsiveness to reasons, hence poised for justification, whether logical thinking (including inference and theory-construction) or practical thinking (including deliberation and decision-making).

According to The Minded Animalism Theory of personhood and personal identity that I work out and defend in *Deep Freedom and Real Persons*, chapters 6-7, and that I briefly spelled out in section 3.2 above, necessarily all real persons are minded animals, but not all minded animals are real persons. Furthermore, necessarily every real person is also a living organism belonging to some species or another,[202] but not every living organism within a species is a minded animal, much less a real person.

Now all sentient animals are fully minded animals, and conversely. But the notion of a minded animal is not *precisely* the same as the notion of a sentient animal, in that some minded animals are not, strictly speaking, *fully* minded animals. Fully minded animals are animals capable of consciousness. Consciousness, in turn, is the subjective experience of a suitably neurobiologically complex *S*-type animal, namely, a living organism within a species. Consciousness is "subjective" because it necessarily includes an ego or first person along with a capacity (whether merely first-order or also higher-order) for oriented reflexivity or self-awareness in space and time. I call this first necessary component of consciousness *egocentric centering*. So the subjective aspect of consciousness is that it is *egocentrically centered*.

Consciousness is also "experience," however, because it necessarily includes both representational content ("intentional content") as well as primitive bodily awareness and other sensations, emotions, feelings, and affects—particularly desires, and pleasure or pain—along with their specific phenomenal content ("phenomenal character"). I will call this second necessary component of consciousness *contentfulnes*s, where this notion is broad enough to include both intentional content and phenomenal character. So the experiential aspect of consciousness is that it is *filled with content*.

In this way, fully minded animals—namely, sentient animals—are subjectively experiencing animals, animals with egocentric centering and contentfulness, hence animals capable of consciousness. For many theoretical purposes, the notions of consciousness, subjectivity, experience, and sentience can all be treated as necessarily equivalent. But as I have defined these notions, experience is not *precisely* the same as consciousness, since it seems clear enough that not every living creature capable of having experiences of some sort or another is also capable of having specifically *subjective* experiences, egocentrically-centered episodes with representational content and phenomenal character.[203] For example, it is plausible to hold that "simple minded" creatures like cephalopods, fish, insects, reptiles, and other non-vertebrates have at least proto-sentience, that is, a capacity for

experiential, contentful, episodes of some minimal sort, yet lack egocentrically-centered mental acts or states.

By the "proto-sentience" of a "simple minded" animal, then, I mean a living creature's non-mechanical responsiveness to external stimuli, together with some proprioceptive capacity, some capacity to have desires, and some capacity to feel pleasure and pain. Although they clearly have proto-sentience, nevertheless cephalopods (for example, octopuses), fish (for example, salmon), insects (for example, mosquitoes), reptiles (for example, snakes), and other non-vertebrates all just as clearly lack the capacity for consciousness—unlike bats, bears, birds, cats, cows, dogs, horses, lions, mice, sheep, and wolves, who just as clearly have a capacity for consciousness and thereby share with us one of our constitutively necessary psychological capacities, "sentience full-stop," as it were.

In this way, proto-sentient, simple minded animals like cephalopods, fish, insects, reptiles, and other invertebrates are certainly *neither non-minded animals*—like, for example, amoebas, human zygotes, human infants with anencephaly, or human adults in a persistent vegetative state—*nor zombies in the philosophical sense*.[204] But at the same time the proto-sentient, simple minded animals are also *not*, strictly speaking, conscious, sentient, or fully minded. They also possess the minimal rudiments of minded animal agency, and thereby are proto-agents, capable of carrying out non-determined, non-indeterministic, non-mechanized, teleologically-driven, spontaneous, actively guided intentional body movements.[205]

Now according to what I have called "The Deep Consciousness Thesis,"[206] *any* sort of mentality or mindedness whatsoever includes at least a minimal degree of occurrent consciousness, which in turn entails at least a minimal degree of occurrent sentience. Therefore proto-sentient, simple minded animals are capable of *some sort of experience*, although they are *not* capable of subjective experience per se. Otherwise put, they have *some* psychological abilities or dispositions that effectively operate when appropriately triggered, which collectively do indeed add up to some kind of animal mindedness, although they do *not* have the capacity for consciousness, or for any other capacity grounded on the capacity for consciousness, per se. A fascinating example is the octopus, a simple minded animal whose proto-sentient mind is almost *literally* spread out all over its body—insofar as its body is almost entirely *arms*, and the majority of the neurons in its body exist outside its brain.[207]

This distinction between simple minded animals and fully minded animals, and correspondingly, the distinctions between proto-sentience and sentience, and between proto-agency and agency, are all directly relevant to the distinction between non-persons and real persons, because they collectively tell us something crucial about the relation between non-persons and moral value. Real persons, as we know, are primary subjects of dignity and primary targets of respect. Sentient, fully minded non-person non-human animals are primary subjects of moral value and targets of moral concern—for example,

bats, bears, birds, cats, cows, dogs, horses, lions, mice, sheep, and wolves. But the scope of moral value and moral concern also extends somewhat *beyond* sentient or fully-minded non-person non-human animals to proto-sentient, simple minded non-human animals—for example, cephalopods, fish, insects, reptiles, and other invertebrates. Proto-sentient, simple minded non-person non-human animals are all at the very least *experiencers* of moral value and targets of moral concern. In other words, even proto-sentience and simple mindedness in animals still matters morally, beyond the limits of real personhood and the capacity for consciousness.

But *why* does even proto-sentience and simple mindedness in animals matter morally? The correct answer to that question, I believe, lies in a direct philosophical appeal to *the capacity to experience pain*, such that *pain* is the direct, intimate, and endogenous (and, in the case of sentient, fully minded animals, reflexive or self-referring) witness to the fact that a minded animal, whether proto-sentient and simple minded or sentient and fully minded, is being harmed. This brings us up to the morally fundamental topics of pain and *suffering*.

4.4 PAIN AND SUFFERING

The available psychological, neurobiological, and ethological evidence strongly suggests that all minded animals, whether human or non-human, experience pain.[208] Now what medical researchers and cognitive neuroscientists call *nociception* is the neurobiological process underlying the experience of pain in suitably complex living organisms. By contrast, what I will neologistically call "nociperception" is the *experience* of pain in minded animals. Given The Deep Consciousness Thesis, and given the notion of minimal sentience, all nociception entails at least minimal nociperception.

Nociperception, as I am understanding this notion, is the experience of a minded animal in direct response to tissue damage or neurobiological systemic disruption caused by various intrusive exogenous stimuli such as burns, cuts, and collisions, or by various noxious endogenous stimuli including relatively enduring conditions such as disease or neurosis, and more temporary conditions such as migraine or emotional distress. Or in other words, nociperception is the direct, intimate, and endogenous—and, in the case of sentient, fully minded animals, reflexive or self-referring—witness to the fact that a minded animal is being harmed. Sentient, fully minded nociperception also includes an egocentric centering of pain in the essentially embodied conscious animal subject. But even proto-sentient or simple minded nociperception includes *some* sort of non-centered, or relatively unfocused, feeling of pain.

In sentient, fully minded animals, nociperception is almost always, but not necessarily—for reasons we shall see shortly—something that the subject of pain-experience does not want. The normal unwantedness of experienced pain is not surprising,

however. The overall function of nociperception is to detect exogenous or endogenous damage, disruption, or distress, and thereby to witness the fact of harm to the living organism: hence the subjective experience of pain directly, intimately, and reflexively witnesses the fact that something bad is happening to the minded animal. Nevertheless this function can be disrupted. For example, in the relatively rare case of congentital insensitivity to pain with anhidrosis, or CIPA, it is possible for subjects to be exogenously damaged and harmed without actually having the subjective experience of pain, or being in nociperceptual states.[209] It is also worth noticing, however, that subjects with CIPA can also subjectively experience endogenously-generated nociperception, for example, emotional distress, hence CIPA is consistent with the general thesis that all minded animals experience pain.

It is empirically known that both the degree and also the specific character of nociperception are not wholly determined by the amount of tissue damage or neurobiological systemic disruption, but instead, in self-conscious or self-reflective, hence rational, human minded animals at least, are partially determined by other factors such as anxiety-level, attention, prior experience, and suggestion.[210] This is what I will call *the subject-dependency of self-conscious pain*.

Moreover it is widely held by contemporary philosophers of mind (although it is of course not wholly uncontroversial) that the causal-functional characterization of pain—that is, pain as characterized abstractly and relationally in terms of the overall pattern of causal transitions from sensory and behavioral stimulus inputs to the minded animal, through the specific neurobiological constitution of the animal, to behavioral outputs from the minded animal—can be held fixed, while systematically varying, across the actual world as well as across logically and metaphysically possible worlds, the specific neurobiological constitution of the minded animal.[211] This is what I will call *the multiple realizability of pain*.

Finally, it is also widely held by contemporary philosophers of mind (although again it is of course not wholly uncontroversial) that the specific phenomenal character of the subjective experience of pain can be held fixed, while systematically varying, again across the actual world as well as across logically and metaphysically possible worlds, the causal-functional characterization of pain.[212] This is what I will call *the multiple functionality of pain*.

Nociperception, whether proto-sentient/simple minded, sentient/fully minded, or self-conscious/self-reflective, also needs to be distinguished from *pain-behavior*. Behavior in general is how a minded animal moves or orients its own living body, or at least is disposed to move or orient its own living body, in response to exogenous or endogenous stimuli. Pain-behavior in particular is the characteristic set of dispositions for unlearned or uncultivated species-specific behavioral responses to tissue damage or neurobiological systemic disruption. Correspondingly, it seems clearly correct to hold that necessarily, other things being equal, if a minded animal is "in" pain, or experiencing pain—that is, is

in a nociperceptual state—then it will also either occurrently exhibit or at least be disposed to exhibit the characteristic pain-behavior of its species.[213]

This necessary truth about pain-behavior also implies, in turn, that pain-behavior is, under some special enabling conditions, a reliable indicator of nociperception. But nociperception is not, strictly speaking, identical to pain-behavior; nor is pain-behavior a strictly sufficient condition of nociperception. This is because when other things are not equal and the special enabling conditions are not satisfied, it is logically and also really possible for a minded animal or indeed even an entire species of minded animals, to be in pain, or experience pain, and thereby be in nociperceptual states, but fail occurrently to exhibit, or even to be disposed to exhibit, the characteristic pain-behavior of its species.[214] Conversely, even when other things are equal, it remains logically and also really possible for all the members of that species to *fake* pain by exhibiting the relevant pain-behavior, without actually also experiencing pain or being in a nociperceptual state. So pain-*behaviorism* is false.

This brings me to a crucial distinction between

(i) *bodily nociperception*, the experience of pain in a minded animal's own living body, and
(ii) *suffering*, self-conscious or self-reflective emotional pain.

Bodily nociperception is pain-experience that is phenomenologically spatially localized in some part or parts of the minded animal's own living body for a certain definite duration of time, and also has a bodily cause. The actual bodily cause of bodily nociperception might not be spatially localized in the same area in which that bodily nociperception is phenomenologically spatially localized. This is vividly evident, for example, when the experience of pain is phenomenologically spatially localized in a "phantom limb." Nevertheless, bodily nociperception is necessarily always phenomenologically spatially localized somewhere or another in the minded animal's own living body, or, in the case of self-conscious/self-reflective minded animals like rational human minded animals, in its body-image.[215] Furthermore, all bodily nociperception has a bodily cause, in the sense that some event inside or at the surface of the minded animal's living body is a sufficient condition, under some psychological-cum-neurobiological law—and in particular, a law that is "hedged" or ceteris paribus[216]—of an episode of pain-experience that is not earlier than the first event. Nevertheless, given the subject-dependency of conscious pain, the degree of nociperception may be altogether out of proportion to the actual extent of tissue damage or neurobiological systemic disruption.

Bodily nociperception is also multiply realizable and multiply functional. For example, it is both logically and also really possible for human minded animals and bats to have the same causally-functionally characterized type of bodily nociperception, by being burned or cut; and it is also logically and really possible for the same subjective experience of

bodily pain to have different causal-functional characterizations and thereby play different causal-functional roles, for example, the ordinary subjective experience of being burned or cut *vs.* masochism.

Suffering, by contrast, is the self-conscious or self-reflective emotional pain of a rational minded animal or real person, which may or may not also involve any bodily nociperception. Suffering therefore need not necessarily be spatially phenomenologically localized—although suffering is always experienced during a certain definite duration of time—and need not necessarily have a bodily cause that is also the cause of bodily pain. Like bodily nociperception in self-conscious/self-reflective minded animals, suffering too is

> (i) *subject-dependent*, which means that the degree or specific character of suffering is partially dependent on anxiety-level, attention, prior experience, and suggestion,
> (ii) *multiply realizable*, since Great apes, other primates, perhaps dolphins, and conceivably Martians—or, slightly closer to home, species-wise, the "Nexus VI replicants" of Ridley Scott's classic 1982 science fiction film *Bladerunner*—can suffer too, and
> (iii) *multiply functional*, since the same subjective experience of suffering can play different causal-functional roles: for example, there are suffering-masochists, just as there are bodily-pain-masochists.

Moreover, both bodily nociperception and suffering can at least sometimes be alleviated by drugs: sometimes by the same drug (for example, alcohol), although usually by different ones (for example, ibuprofen *vs.* anti-depressants). Yet both suffering and bodily nociperception cannot *always* be alleviated by drugs. Certain kinds of emotional pain are remarkably resistant to pharmacological remedy, and there are also certain kinds of awful "central" bodily nociperception that are similarly drug-resistant—although they may still respond to neurosurgery.

In real human persons like us, bodily nociperception can materially constitute suffering, which is to say that in higher-level or Kantian rational minded human animals there can be a spatiotemporal coincidence and also a metaphysical dependence relation—more specifically, a grounding relation—between the subject's phenomenologically localized self-conscious/self-reflective bodily nociperception and the cause of suffering. For example, I can suffer when my right leg hurts and just because my right leg has been damaged. But in principle, that token experience of suffering could have been spatiotemporally coincident with another different phenomenologically localized bodily nociperception, and similarly with a different token of that type of suffering: I might have identically suffered, whether it was my left leg or my right leg that was hurting (so token suffering can be preserved under phenomenological enantiomorphism in the animal body); and I might have suffered in just the same way, whether it was my leg or arm or head that was hurting (so type suffering can be preserved under change of phenomenological spatial localization in the animal body).

This points up the absolutely crucial fact that bodily nociperception is not equivalent with suffering, despite the obvious fact that bodily nociperception and suffering very often go together.

It is possible to experience mild or intense bodily nociperception but not suffer at all—for example, high performance athletes, women during childbirth, professional dancers, solitary masochists, consensual sadomasochistic sex-partners, ancient Greek Stoics, and so-on. Indeed, for at least some of these people, at least some of the time, the specific character of the experience of pain is positively and self-consciously needed or wanted.

Conversely, it is also possible to suffer mildly or intensely but not be in any sort of bodily nociperceptual state—for example, extreme embarrassment, extreme shyness, guilt, jealousy, extreme disappointment, anxiety, fear, depression, and also the kind of suffering that is the result of certain forms of emotional trauma such as rejection, betrayal, loss of a loved one, etc.—which might be collectively called *grief*.

It is of course true that extreme embarrassment, extreme shyness, guilt, jealousy, extreme disappointment, anxiety, fear, depression, and grief can also be combined with bodily nociperception, just as they are usually combined with bodily reactions like flushing, heart palpitations, shivers, turning pale, or sweating. My point is simply that the various forms of suffering are not strictly always or necessarily combined with the experience of bodily pain, just as they need not strictly always or necessarily be combined with flushing, heart palpitations, shivers, turning pale, or sweating.

Does the non-equivalence of bodily nociperception and suffering still seem implausible to you? Let me try another line of argument.

The same basic point can also be indirectly made by recalling the necessary ceteris paribus connection between being in pain or experiencing pain and pain-behavior: necessarily, other things being equal, if a minded animal is in pain or experiences pain, then it will also exhibit the pain-behavior of its species. Now think of, or imagine, what a human minded animal who is in a considerably intense nociperceptual state and also intensely suffering looks like and behaves like—for example, someone being tortured—and then contrast that with what a human minded animal who is in a relevantly similar sort of considerably intense nociperceptual state and yet not suffering at all looks like and behaves like—for example, long distance runners or consensual sadomasochistic sex-partners. Correspondingly, again think of, or imagine, what a human minded animal who is in a considerably intense nociperceptual state and also intensely suffering looks like and behaves like—for example, again, someone being tortured—and then contrast that with what the behavior of a human minded animal who is not in any sort of bodily pain and yet also intensely suffering to a relevantly similar degree looks like and behaves like—for example, someone wracked with guilt.

In each pair of cases, I think, the manifest visual and behavioral differences are radically sharp. This in turn strongly suggests that the conceptual distinction between bodily nociperception and suffering is something that we all clearly recognize.

So far I have been using, as a preliminary general characterization of suffering, that it is the self-conscious or self-reflective emotional pain of a rational minded animal or real person. But what *more specifically* is suffering?

My Existential Kantian Ethics-based view is that, as opposed to mere bodily nociperception, *suffering essentially expresses a self-conscious or self-reflective rational minded animal's direct, intimate, endogenous, and reflexive sense of harm to the constitution of its own will and to its own real personhood*. All experience of pain directly, intimately, and endogenously witnesses a minded animal's being harmed. But, more specifically, real personal suffering directly, intimately, and reflexively witnesses a self-conscious or self-reflective real person's being harmed in its own capacity for intentional agency—hence "where it really hurts" or "right where she lives"—that might or might not also involve bodily harm.

In other words, my suffering witnesses the fact that I am being harmed *in respect of what I self-consciously or self-reflectively care about and want most deeply*. To suffer is to subjectively experience emotional pain *for a practical reason*.

This is not to say that I always self-consciously or self-reflectively represent that practical reason to myself, that I am always at that time choosing or acting on that reason, or even that, recognizing that reason, I am always prepared to adopt it henceforth as mine. Indeed, the cause of suffering is very often brutally or brutely imposed upon us, altogether against our wills, and without self-consciousness or self-reflection, for example, in the sufferings of those in the grip of a non-catastrophic but irremediable mental illness. And suffering is rarely, if ever, the result of self-conscious or self-reflective deliberation and future planning—although it seems at least barely conceivable, but perhaps no more than that, that someone could plan to suffer.

Nevertheless, it remains true that at least in principle *I can understand the reason why I am suffering*. More generally, it remains true of all self-conscious and self-reflective higher-level or Kantian real persons like us, that there is always a certain minimal sense in which *we choose to suffer*, in that suffering is always both motivated and justified by reasons which, at least in principle, we can become self-consciously and self-reflectively aware of, which we can understand, and which we can self-consciously and self-reflectively either adopt or reject as our own.

This admittedly unusual line of thinking leads to a perhaps surprising conclusion. Other things being equal, people are neither blamed nor praised by others for their suffering: instead, other things being equal, they are only pitied for their suffering. But at the same time, if I am a self-conscious and self-reflective, autonomous, rational minded animal, a higher-level or Kantian real person, then I am still in a certain minimal sense *deeply (non-)morally responsible for my own suffering*. I am the *ultimate source* of it, and it is *up to me*. Anyone's suffering may well be, and very often is, *absolutely not his or her fault*. So the point I am making is *absolutely different from "blaming the victim."* Nevertheless, in the special sense I have just described, my suffering is "my thing," my Sisyphean rock to

push all the way up that cursed hill, temporarily, or day after day, endlessly. Suffering expresses my direct, intimate, endogenous, and reflexive sense of harm to my own rational human minded animal capacity for intentional agency.

Therefore, at least in principle, I can become self-consciously and self-reflectively aware of the practical reason why I am suffering, and get a rational and emotional handle on it, and then self-consciously and self-reflectively *either accept it or refuse it*. So, ultimately, in such cases, I can *either suffer or not suffer*, either by freely accepting this harm as harm-to-me, or by freely refusing this harm as harm-to-me. Especially in the case of refusing-to-suffer, I am not saying that this is in any way easy to do. In fact, it may be fantastically difficult to do. But I do think that it is at least really possible for a higher-level or Kantian real human person to do. As per Wittgenstein's Mystical Compatibilism in the *Tractatus*, even while the physical facts all remain the same, I can make the normative structure of my world wax or wane by resolutely turning it into *the world of the happy*. This is also what Kant calls a *revolution of the heart* or *revolution of the will* (*Rel* 6: 48).

So, for all these reasons, the somewhat clichéd saying, "pain is inevitable, but suffering is optional," turns out to have more than a grain of existential and moral truth in it.

It is crucial to note, however, that being able *to understand* a practical reason for one's own suffering, and *being deeply (non-)morally responsible* for one own suffering, are not characteristic features of lower-level or Frankfurtian real persons, such as normal, healthy toddlers or other normal, healthy older children. For better or worse, these lower-level real persons can suffer *for* a practical reason without in any way being able *to understand the reason why*, simply because they lack the sophisticated conceptual competence and the sophisticated reflective capacity to do so, and thus they are not in any way responsible for their own suffering, even in the special sense I have been spelling out. Lower-level or Frankfurtian real persons are self-*conscious* without also being self-*reflective*.

In any case, once we recognize that suffering is a higher-level or Kantian real human person's self-conscious or self-reflective experience of emotional pain for a practical reason, then I think that we can also recognize that suffering falls naturally into three distinct categories, corresponding to the three main categories of things that practical reasons can be about

(i) my practical relations to myself,
(ii) my practical relations to the world, or
(iii) my practical relations to other real persons.

Thus it seems clear that

either (i*) I suffer because I am not what I want to be or the way I want to be (self-emanating suffering),
or (ii*) I suffer because the world is not what I want it to be or the way I want it to be (world-emanating suffering),

> or (iii*) I suffer because other real persons are not what I want them to be or the way I want them to be (other-emanating suffering).

This tripartite scheme, in turn, carries over aptly and smoothly into folk wisdom about the nature of suffering. I am "my own worst enemy" (self-emanating suffering). The world is "a vale of tears" (world-emanating suffering). And "hell is other people" (other-emanating suffering).

Of course, and sadly, suffering can also involve combinations of the three basic kinds. For example, if someone I deeply love dies, I might suffer as much from thinking that I miserably failed to treat her as lovingly as I should have (self-emanating suffering), as I also do from the truly awful fact that she is simply no longer there in the world to be with me (other-emanating suffering); and at the same time, I may also suffer equally intensely from the thought that her permanent absence from the world and from my life is nothing but a cruel joke (world-emanating suffering).[217]

So to summarize:

> Suffering is the self-conscious or self-reflective experience of emotional pain—anguish, despair, grief, sorrow, and so-on—consequent upon a real person's being a direct, intimate, endogenous, and reflexive witness to her being harmed in her capacity for intentional agency, and in what she cares about and wants most, either by her own means, by means of the world, by means of other people—or, again sadly, also by means of any two or three of the above.

It seems self-evident, given the nature of suffering, that the suffering of any real person is always a target of serious moral concern for every rational animal or real person. Otherwise put, wherever and whenever suffering happens, it is never morally insignificant or irrelevant. It does not follow, however, *that we are morally obligated to prevent or reduce suffering in ourselves or others*—that is morally great, and morally heroic, but also supererogatory. But it does follow that we are morally obligated *always to heed suffering and to take it into account in our moral choices, actions, and deliberations, other things being equal*. There is, however, one important qualification to this relatively weak moral principle that I will spell out four paragraphs below, according to which we are morally obligated, other things being equal, never intentionally to cause, and also always to prevent or reduce, a certain special kind of suffering in real persons.

In view of The Minded Animalism Theory of personhood and personal identity, together with the Existential Kantian Ethics-based account of suffering as a real person's self-conscious or self-reflective experience of emotional pain for a practical reason, then it directly follows *that we always have a sufficient practical reason to fear our own future suffering*. Significantly, neither any the classical or standard contemporary approaches approaches to personal identity—especially including Parfit's—nor ethical egoism, nor act consequentialism, whether they are combined with any of these classical or standard

contemporary approaches to personal identity or not, can adequately explain or justify this highly plausible thesis,[218] a philosophical fact which is nowadays called *The Non-Identity Problem*. Psychological criteria for personal identity fail, because of cases in which bodily continuity is preserved over a break in psychological continuity. Bodily criteria for personal identity fail, because of cases in which psychological continuity is preserved over a break in bodily continuity. And the Parfitian account also fails, because "identity is not what matters." Ethical egoism fails too, because it always privileges my emotional states in the present moment over those in the future. And act consequentialism also fails, because it always privileges everyone else's aggregated future emotional states over mine alone. Therefore, the fact that Existential Kantian Ethics can smoothly explain and justify the highly plausible thesis that we always have a sufficient practical reason to fear our own future suffering, and also the further fact that Existential Kantian Ethics thereby arguably solves The Non-Identity Problem, collectively deliver a decisively important philosophical test-case result in support of the view I am developing in this book and in this chapter.

But most importantly for the present purposes of my argument, we can now see *that suffering can occur without bodily nociperception*, and also *that bodily nociperception can occur without suffering*. So bodily nociperception does not strictly always or necessarily involve suffering. This fact in turn directly implies that Bentham's and Singer's famous direct inference from the mere fact of a minded animal's experience of bodily pain to that minded animal's suffering is fallacious: for later reference, I will call this *The Bentham-Singer Fallacy*.

The recognition of The Bentham-Singer Fallacy also leads on to a deeper point. The crucial feature of suffering, as opposed to bodily nociperception, is that necessarily, all and only real persons can suffer. That is because suffering requires a psychological complexity, in virtue of constitutively necessary capacities, that is characteristic of all and only real persons, whether human or not. In other words, as real persons, whether human or not, we are essentially the only conscious animals in the universe who can *suffer*—and choose and do *evil*, whether banal evil, like Arendt's Eichmann, under the guise of the good, or near-satanic evil, like Shakespeare's fictional Iago, Cormack McCarthy's fictional Anton Chiurgh, or real-world Hitler, under the guise of the bad.[219]

Lucky us.

But from this it also directly follows that non-persons *cannot* suffer, even if they *can* subjectively experience bodily pain or have episodes of conscious bodily nociperception, and therefore even if they are sentient, fully minded animals.

This has fundamental moral implications. Treating a real person with respect means never intentionally harming that real person by violating her dignity, other things being equal. Such a violation causes a certain special kind of suffering in the real person whose dignity has been violated—the suffering of someone *who has been harmed by being disrespected and treated as a mere thing*. Let us call it *degradation*.

Degradation can be contrasted with *oppression*, which is treating people in ways that fall saliently below what is minimally sufficient to meet the moral demands of respect for their human dignity. It is possible to be oppressed without being degraded—for example, the very fact of human persons working in shit jobs as wage-slaves, living in poverty, or without adequate healthcare, anywhere, when the means to alleviate the worst effects of big-capitalism, to end poverty, or to provide adequate healthcare are available—is always oppression but not always degradation. This is because oppression can occur through *institutional* means, not only in past, recent, or contemporary authoritarian or totalitarian states, but also in contemporary (neo)liberal democratic states, without any one individual's explicit or self-conscious intention to harm.[220]

In any case, all degradation is *also* oppression, because degradation involves coercion, and coercion is a necessary and minimally sufficient condition of oppression.[221]

Moreover and above all, as I think we all know with self-evidence, it is truly awful for someone to be degraded by someone else. Degradation may of course also literally kill the degraded real person. But even if degradation does not literally kill you, nothing in this world is subjectively or objectively worse than someone's treating you like a mere thing, like a piece of garbage or offal, fully without your actual or possible rational consent, and with manifest cruelty. And for the degraded one, that is, in some cases *you*, being treated this way burns and hurts like the fires of hell, only it is even more terrible, since it is utterly undeserved.

This is one central reason why, for example, torture, lynchings, rape, and child abuse or child murder are such heinous moral crimes: they are so patently treating someone like a mere thing, coercively, without their consent, and cruelly. So sufficiently treating a real human person with respect for their dignity also means being morally obligated *never* to treat that person in such a way as intentionally to cause them *that* kind of suffering, degradation, and also *always* in such a way as to prevent or reduce degradation, other things being equal. Therefore, and now fully explicitly, other things being equal, we are morally obligated *never* to treat real human persons, and also any non-human real persons there might be, in such a way as to degrade them, and also *always* in such a way as to prevent or reduce their degradation.

At the same time, however, it is *not* the case that we are morally obligated generally to prevent or reduce the bodily nociperception of real persons, whether human or non-human, or even morally obligated generally to prevent or reduce the bodily nociperception of sentient, fully minded animals, for two reasons.

First, in real persons, whether human or non-human, *not* all bodily nociperception entails suffering, and *not* all suffering entails bodily nociperception. Therefore, other things being equal, we are not morally obligated to prevent or reduce bodily nociperception in those real human persons like us who are *capable of* suffering, but are not *actually* suffering, even though they are undergoing bodily nociperception. For example, other things being equal, we are *not* morally obligated to prevent or reduce anyone's everyday

experience of minor bodily pains due to bumps, eye-strain headaches, hangnails, scratches, sore muscles after exercising, stubbing one's toe, and so-on. *Nor*, other things being equal, are we morally obligated to prevent or reduce the experience of even intense bodily pain in high performance athletes, women during childbirth when they have specifically chosen natural birthing, professional dancers, solitary masochists, consensual sadomasochistic sex-partners, or Stoics, even if there would be some moral value in doing so, and even if it would also be morally *permissible* to do so.

The case of bodily nociperception in women during childbirth when they have specifically chosen natural birthing, is a particularly good example of this. In such cases, because of their rationally-formed personal views about natural childbirthing and their high levels of commitment to this project, these women self-consciously prefer and specifically request in advance that *no* experience of bodily pain during childbirth, even highly intense bodily pain, be prevented or reduced by the use of medical anaesthetics.[222] That is perfectly reasonable and morally permissible. By sharp contrast, many other pregnant women self-consciously prefer and specifically request that *all* experience of bodily pain be prevented or reduced by the use of medical anaesthetics. And that is perfectly reasonable and morally permissible too. Furthermore, some *other* pregnant women, belonging to neither the pro-natural-birthing-pain group or the anti-birthing-pain group, self-consciously prefer and request *waiting to find out* what their bodily nociperceptive pain-levels are actually like during childbirth itself, *then* request or refuse medical anaesthesia: and that is also perfectly reasonable and morally permissible.

Therefore, bodily nociperception in real human persons like us, even highly intense bodily nociperception, other things being equal, is morally neutral.

Second, some animals that are capable of bodily nociperception—whether proto-sentient and simple minded animals or sentient and fully minded animals—are *not* capable of suffering, precisely because they are not real persons. So obviously, then, we are not morally obligated generally to prevent or reduce the bodily nociperception of *non*-person minded animals, *if* we are not morally obligated generally to prevent or reduce the bodily nociperception of *rational* minded animals or real persons.

Does this mean that non-rational sentient and fully minded animals in bodily pain, or even proto-sentient and simple minded animals in bodily pain, may be treated like mere things?

No. These animals are still experiencers of moral value and primary targets of our moral concern—that is, they have *interests*, and we must heed those interests—and we are therefore obligated, at the very least, to consider them, and to take them fully into account in our moral reasoning. In the case of sentient, fully minded animals, this involves treating them as primary, *serious* targets of our moral concern. And in section 4.6, I will examine more precisely what it means to treat sentient, fully minded animals as primary, serious targets of our moral concern. But in the next section I will consider the *moral comparison* between, on the one hand, our obligations to the minded animals who can suffer, and, on

the other hand, our obligations to the minded animals who cannot suffer but instead are capable only of experiencing bodily pain.

4.3 MORAL COMPARISON

In this section, I want to argue for what in section 4.2 I called *The Moral Comparison Thesis*, which says:

> There is compellingly good reason to believe that the suffering of any human or non-human minded animal that *is* a real person, whether via bodily nociperception or without bodily nociperception, is substantially more morally significant than the bodily nociperception of any human or non-human minded animal that is *not* a real person, assuming roughly comparable levels of experienced intensity.

In order to carry out that argument, I will also need to define some terminology. By "moral significance," in the present context, I mean the following:

> *X* is morally significant if and only if *X* has moral value, and the presence or absence of *X* in the life of an experiencer *E* not only makes a determinate, noticeable, life-modulating difference in the life of *E*, but also partially determines the application of moral principles to *E*.

In turn, by "life-modulating difference," I mean a difference that saliently affects the content or course of one's life (for example, starting a romantic relationship, ending a romantic relationship, moving to another city or country, losing your job, starting a new job, etc.), without necessarily implying a life-changing difference in the stronger sense of radically restructuring the content or course of one's life (for example, falling permanently in love, experiencing the death of a loved one, falling into the grip of a serious addiction, mastering a serious addiction, finding one's permanent calling or vocation in a non-religious sense, religious conversion, etc.).

Given those definitions, my argument for The Moral Comparison Thesis will deploy five basic premises:

> (1) *The Mental-Mental Gap Thesis*, originally defended by Nagel,
> (2) the sharp distinction between *bodily nociperception* and *suffering*, that I argued for in section 4.3, which entails that it is fallacious for Bentham and Singer to infer the existence of suffering from the mere fact of a minded animal's experience of bodily pain,
> (3) *The Multiple Realization Thesis*, originally defended by Putnam,
> (4) *The Structure-Restricted Correlation Thesis*, originally defended by Kim, and also something I call

(5) The Schematization Thesis.

In the rest of this section, I will unpack and offer justification for premises (1), (3), (4), and (5), and then explicitly lay out my argument for The Moral Comparison Thesis.

(1) The Mental-Mental Gap

Nagel's classic essay, "What Is It Like To Be A Bat?," is all about *explanatory gaps*.[223] An explanatory gap obtains just in case one set of concepts cannot be reduced to or entailed by another set of concepts, whether by analytical definition, analytical entailment, or even by some weaker kind of reduction such as necessary coextension. If every explanatorily irreducible set of concepts picks out a set of distinct properties and facts in the world, then every explanatory gap entails a corresponding *ontological gap* and failures of logical supervenience or nomological supervenience at the level of properties and facts. The inferential step from explanatorily irreducible concepts to distinct properties and facts has been much discussed since the first publication of Nagels's essay in 1974, and remains controversial. What many readers of "What Is It Like To Be A Bat?" over the last 45 years seem *not* to have noticed, however, is that Nagel actually discloses *two* different explanatory gaps in the philosophy of mind and not just *one*.

First and foremost, there is his well-known explanatory gap between mentalistic concepts and physicalistic concepts.[224] Mentalistic concepts are concepts whose content and ascription imply full consciousness or subjective experience, and the first-person point of view, or in Nagel's terms, "what it is like to be, for an organism":

> [F]undamentally an organism has conscious mental states if and only if there is something it is like to *be* that organism—something it is like *for* the organism. We may call this the subjective character of experience.[225]

Subjective character is specific phenomenal character, for example, the quite peculiar feeling experienced by a certain kind of cinephile, of being simultaneously bored stuffless and also intensely saddened, by virtue of watching the truly awful 1959 Rock Hudson – Doris Day movie *Pillow Talk*.[226] So mentalistic concepts are concepts whose content and ascription imply *specific phenomenal character*. Physicalistic concepts, by contrast, are concepts whose content and ascription imply *only first- or second-order physical properties or facts, the exclusively non-subjective and objective character of the natural world, and the third-person/impersonal point of view*: so roughly, *what it is like for something to be fundamentally or superveniently physical*. Nagel's claim is that physicalistic concepts can never adequately capture or explain the specific phenomenal character of subjective experience. Let us call this *The Mental-Physical Gap*.

But second, for Nagel there is also a seemingly equally intractable explanatory gap between the mentalistic concepts that we apply to the conscious states of animals belonging to our own species (aka "conspecific animals"), and the mentalistic concepts that we apply

to the conscious states of animals belonging to other species (aka "heterospecific animals"). Nagel's claim is that although we are capable of understanding the specific phenomenal character of the subjective experience of *con*specific animals (hence there is no general skeptical problem of other minds, at least for other conspecific minded animals), nevertheless we are incapable of understanding the specific phenomenal character of the subjective experience of heterospecific minded animals: for example, what it is like for a bat to get around in the world by echolocation. As he puts it in another essay:

> We ascribe experiences to animals on the basis of their behavior, structure, and circumstances, but we are not just ascribing to them behavior, structure, and circumstances. So what are we saying? The same kind of thing we say of people when we say they have experiences, of course. But here the special relation between first- and third-person ascription is not available as an indication of the subjectivity of the mental. We are left with concepts that are anchored in their application to humans, and that apply to other creatures by a natural extension from the behavioral and contextual criteria that operate in ordinary human cases. This seems definitely unsatisfactory, because the experiences of other creatures are certainly independent of the reach of an analogy with the human case. They have their own reality and their own subjectivity.[227]

I will call this *The Mental-Mental Gap*. One direct implication of The Mental-Mental Gap that Nagel does not explicitly mention, but which will be highly relevant to us later, is what I will call *The Nagel Proportionality Thesis*:

> The greater the degree of neurobiological and behavioral difference between the human species and another minded animal species, then the wider The Mental-Mental Gap between us and those heterospecific animals.[228]

Those closely familiar with the argument-structure of "What Is It Like To Be A Bat?" will also notice that I have reversed Nagel's order of argumentation. In fact, he first argues for the existence of The Mental-Mental Gap, and then, second, he uses that gap as the basic premise in his argument for the Mental-Physical Gap. Officially then, The Mental-Mental Gap is supposed to be the sufficient reason for The Mental-Physical Gap:

> This [Mental-Mental Gap] bears directly on the mind-body problem. For if the facts about experience—facts about what it is like *for* the experiencing organism—are accessible from only one point of view, then it is a mystery how the true character of experiences could be revealed in the physical operation of that organism.[229]

In other words, for Nagel, The Mental-Mental Gap is supposed to *entail* The Mental-Physical gap. But this is a mistake. Although The Mental-Mental Gap is perfectly *consistent* with The Mental-Physical Gap, nevertheless the two gaps are *logically independent*. This is because it is perfectly coherent to hold that we cannot understand the

specific character of the subjective experience of heterospecific minded animals and *also* hold that reductive physicalism is true, hence there is no Mental-Physical Gap.

It is true that at least one version of reductive physicalism—a hyper-strong version of type-type physicalism, or mind-brain identity theory, that implies both the analytically necessary identity of mentalistic concepts with physicalistic concepts, and also the metaphysically necessary identity of mental properties with physical properties—is sufficient for closing The Mental-Mental Gap. But *not every* version of reductive physicalism is of this hyper-strong sort. Indeed, most reductive physicalists, including the defenders of classical type-type physicalism, explicitly reject the hyper-strong analyticity version of the mind-brain identity theory and opt for the metaphysically necessary a posteriori identity of mental properties with physical properties, aka "contingent identity," while also rejecting both the analytically necessary identity of mentalistic concepts with physicalistic concepts and the analytically necessary identity of mental properties with physicalistic concepts, alike.[230] Therefore, it is obvious that closing The Mental-Mental Gap is not generally necessary for the truth of reductive physicalism.[231] Nor, indeed, is closing The Mental-Mental Gap necessary for physicalism of any sort, whether reductive or non-reductive.[232] So Nagel's Mental-Physical Gap is logically independent of his Mental-Mental Gap: the latter is consistent with the denial of the former.

From this point forward, then, I am going to assume the logical independence of the two Gaps, and also that Nagel's argument for the existence of The Mental-Mental Gap, in and of itself, especially including The Nagel Proportionality Thesis, is basically sound.

(2) Multiple Realization, Structure-Restricted Correlation, and Schematization

Both The Multiple Realization Thesis and also The Structure-Restricted Correlation Thesis arise out of philosophical debates about *functionalism* in the philosophy of mind.[233] Functionalism in general holds that minds are not separate substances and that mental properties are not intrinsic non-relational properties of something, but instead that mental properties are identical to *functional properties*: that is, extrinsic relational patterns of occurrent or dispositional causal transition from inputs to outputs,[234] applying to the external and internal states of physical machines or living organisms. Standard examples of functional properties are the properties instantiated by sequences of digital computations in a universal Turing machine, which are the special focus of *computational functionalism*, as well as those properties instantiated by those neurobiological processes in the brains and central nervous systems of animals that are apt to cause behavior, which are the larger focus of *psychofunctionalism*. According to *metaphysical functionalism*, whether reductive or non-reductive, functional properties are second-order physical properties[235] that are strongly supervenient on first-order physical properties (downwardly identical or logically supervenient in the case of reductive functionalism, and naturally or nomologically supervenient in the case of non-reductive functionalism). So metaphysical functionalism holds that mental properties are identical to a certain special sort of second-order physical property that is strongly supervenient on first-order physical properties.

Metaphysical functionalism is committed to the truth of both *The Multiple Realization Thesis* and *The Structure-Restricted Correlation Thesis*.

The Multiple Realization Thesis asserts that one and the same functional property can be instantiated across the actual world and also across logically and really possible worlds in many different physical individuals, types of organism, natural kinds, and compositional stuffs.[236] This is partially verified by the empirical facts that the very same kind of computational software (for example, Microsoft Word 2016, which is what this text is being processed in) can be instantiated in many different sorts of hardware, and that the very same generic type of physiological or neurobiological process (for example, digestion or sleep) can be instantiated in many different species of animals (for example, humans, Great apes and other primates, bats, and cats).

The Structure-Restricted Correlation Thesis, on the other hand, asserts the relativization of the instantiation of mental properties to species-specific physical structure types, or as Kim puts it:

> If anything has mental property *M* at time *t*, there is some physical structure type *T* and physical property *P* such that it is a system of type *T* at *t* and has *P* at *t*, and it holds as a matter of law that all systems of type *T* have *M* at a time just in case they have *P* at a time.[237]

In other words, mental properties occur in animals under specific physical conditions in a lawful way, and this lawful regularity is found across species insofar as they share the same basic neurobiological physical constitution. If mental properties are functional properties, then The Structure-Restricted Correlation thesis is a necessary condition of the multiple realizability of mental properties. So if metaphysical functionalism is true, then The Structure-Restricted Correlation Thesis follows automatically. But even if mental properties are *not* identical to functional properties—but only, for example, strongly supervenient on functional properties—nevertheless The Structure-Restricted Correlation Thesis remains true, because all that it says is that minded animal *X* has a mental property at a time only if *X* has a certain physical structure and there is some lawful connection between instantiations of that mental property and fundamental physical properties of that instantiated physical structure. Indeed, it is important to see that The Structure-Restricted Correlation Thesis is perfectly compatible with various denials of mind-body physicalism, whether the physicalism that is denied is reductive or non-reductive. For even if mental properties are neither identical with nor in any sense strongly supervenient on either first-order or second-order physical properties, there can still be structure-restricted correlations between mental properties and physical properties in minded animals.[238]

How can we isolate a given structure-restricted correlation? I think that the correct answer is the one offered by Kim, namely, that we isolate it by

> (i) finding the causal-functional characterization of that mental property, and then
> (ii) correlating the causal-functional role picked out by that characterization with a certain species-specific neurobiological constitution in minded animals.[239]

Again however, it is crucial to remember that we do not have to identify mental properties with functional properties in order to do this. All we that is required is to have a causal-functional characterization of the relevant mental property. Consider, for example, the causal-functional characterization of the mental property of the experience of bodily pain (aka, bodily nociperception) that I sketched above: Bodily nociperception is the minded animal's experience of tissue damage or neurobiological systemic disruption or distress within its own body; bodily nociperception is subject-dependent (in self-conscious or self-reflective animals), multiply realizable, and multiply functional; and bodily nociperception has a necessary connection (other things being equal) to animal behavior. The causal-functional role of bodily nociperception is multiply realized in humans, Great apes and other primates, bats, cats, and so-on, and therefore determines a set of structure-restricted correlations between the mental property of being in bodily pain and different species-specific neurobiological constitutions.

This point leads to one last concept that we will need before I get to the explicit argument for The Moral Comparison Thesis. This is the concept of a *schematization* of a mental property. The basic idea behind the schematization of a mental property is that heterospecific minded animals will typically subjectively experience the same sorts of things—for example, colors, sounds, or pain—quite differently, precisely because their neurobiological constitutions are quite different, and also in some sort of systematic relation to the precise differences in neurobiological constitution. The mental content of a body-schema, in turn, is essentially non-conceptual.[240] Otherwise put, a schematization *is the essentially non-conceptual way that different types of animal bodies directly and endogenously affect the specific phenomenal character of consciousness in those minded animals*. This, in turn, is also sometimes re-presented conceptually in the form of a distinctive *body-image* that gives a correspondingly distinctive underlying phenomenological spatiotemporal organization to a self-conscious/self-reflective animal's primitive bodily awareness and affects.[241] Finally, and slightly more abstractly put, a mental property is schematized if and only if the specific phenomenal character of the essentially non-conceptual mental content of its instances is regularly and systematically modified and shaped (sometimes also via a conceptualized body-image) by the multiple realizations of its corresponding causal-functional role in different neurobiological substrates under different structure-restricted correlations.

My thesis here, then, is that necessarily, all minded animal consciousness is schematized. This is what I will call *The Schematization Thesis*. Human pain and bat pain are both *pain*, in the sense that they each play the same causal-functional role in the human species (homo sapiens) and the bat species (microchiroptera). But in virtue of The

Schematization Thesis, a bat's subjective experience of bodily pain will be radically different from a human animal's subjective experience of bodily pain, not merely neurobiologically, but also phenomenologically. Indeed it is precisely the notion of schematization, and thus implicitly The Schematization Thesis, that drives the basic intuition that Nagel uses to motivate The Mental-Mental Gap Thesis and The Nagel Proportionality Thesis themselves. We lack any sort of adequate conceptual or theoretical understanding of what it is like to be a bat precisely because the subjective experiences of humans and of bats are schematized radically differently, and precisely because the differing essentially non-conceptual contents of schematization across different species are given directly only to conspecific minded animals.

It does *not* follow from this fact, however, that in order to have correct perceptions or make true judgments about other minded animals, whether human or non-human, one must be standing in some sort of identity relation to their mental states, whether type-identity or token-identity. *Nor* does it follow that this cognition of other minded animals happens fundamentally by means of analogical inference or "theory-of-mind." Instead, on my view, correctly perceiving and judging the mental states of other minded animals occurs only by means of pre-reflectively consciously simulating their essentially embodied mental states in oneself, that is, by means of what I call *empathic mirroring*.[242] Empathic mirroring, in turn, is a matter of emotional cognition, not theoretical cognition, and it is mediated by essentially non-conceptual content,[243] not by conceptual content. Empathic mirroring is as effective for cognizing non-human minded animals[244] as it is for cognizing other human minded animals. So in general, we cannot "read" or conceptually understand other minds. But we can, to a greater or lesser extent, as it were, "dance" with, *essentially non-conceptually resonate with*, other minded animals, whether human or non-human. At the same time, where other species are concerned, we must also, to borrow a phrase from The London Underground, "mind the gap"—by which I mean that we must also explicitly accept the existence of Nagel's Mental-Mental Gap.

In any case, here is my explicit argument for The Moral Comparison Thesis, laid out step-by-step.

An Argument for the Moral Comparison Thesis

(1) Assume the notion of a minded animal's consciousness as a capacity for subjective experience, also characterized by Nagel as "what it is like to be, for an organism." (Premise, justified by arguments already provided.)

(2) Assume both The Mental-Mental Gap Thesis and The Nagel Proportionality Thesis: We are incapable of understanding the specific phenomenal character of the subjective experience of heterospecific animals (The Mental-Mental Gap Thesis), and the greater the degree of neurobiological and behavioral difference between the human species and another minded animal species, the wider the Mental-Mental Gap between us and them (The Nagel Proportionality Thesis). (Premise, justified by arguments already provided.)

(3) Assume The Multiple Realization Thesis: One and the same functional property can be instantiated across the actual world and also across logically and really possible worlds in many different physical individuals, types of organism, natural kinds, and compositional stuffs. (Premise, justified by arguments already provided.)
(4) Assume The Structure-Restricted Correlation Thesis: Mental properties occur in animals under specific physical conditions in a lawful way, and this lawful regularity is found across species sharing the same neurobiological physical constitution. (Premise, justified by arguments already provided.)
(5) Assume The Schematization Thesis: Necessarily, all minded animal consciousness is schematized, and a mental property is schematized just in case the specific phenomenal character of the essentially non-conceptual content of its instances is regularly and systematically modified and shaped (sometimes also via a conceptualized body-image) by the multiple realizations of its corresponding causal-functional role in different neurobiological substrates under different structure-restricted correlations. (Premise, justified by arguments already provided.)
(6) Assume that conscious states like being in bodily pain, that is, states of bodily nociperception, have corresponding causal-functional characterizations that describe their causal-functional roles. (Premise, justified by arguments already provided.)
(7) Therefore, the causal-functional role of bodily nociperception is multiply realized under different species-specific structure-restricted correlations. (From (3), (4), and (6).)
(8) Therefore, The Schematization Thesis applies to bodily nociperception in all minded animals. (From (5) and (7).)
(9) Therefore, there is good reason to believe that the bodily nociperception of bats and other neurobiologically heterospecific animals is not only phenomenologically radically different from the human experience of bodily pain but also in fact conceptually inaccessible to us. (From (1), (2) and (8).)
(10) We have good reason to believe that a type of bodily nociperception in heterospecific animals can be anywhere near as morally significant as the bodily nociperception of real human persons only if it is phenomenologically very similar to bodily nociperception in real human persons. This is because our reason for believing that the experience of bodily pain in our own species is morally significant is necessarily based on first-person evidence. But the less phenomenologically similar a given consciousness-type C_1 (say, a bat's bodily nociperception) is to another consciousness-type C_2 (say, a conscious human animal's bodily nociperception), then the less reason we have to believe that C_1 has all or even any of the consciousness-based properties that C_2 has. (Premise.)
(11) Real human persons can suffer, whether via their bodily nociperception or without bodily nociperception, and this suffering is substantially morally significant. (Premise, justified by arguments already provided.)
(12) Therefore, there is compellingly good reason to believe that the suffering of any human or non-human minded animal that *is* a real person, whether via bodily nociperception or without bodily nociperception, is substantially *more* morally significant than the bodily nociperception of any human or non-human minded animal that is *not* a real person, assuming roughly comparable levels of experienced[245] intensity. In other words, The Moral Comparison Thesis is true. (From (9), (10), and (11).)

4.6 KINDNESS TO ANIMALS REVISITED: HARMING WITHOUT TORTURE OR CRUELTY

How ought we to treat non-human minded animals who are not real persons? I have just argued that we have compellingly good reason to believe that bodily nociperception in non-human non-person minded animals is substantially less morally significant than roughly comparable intensity-levels of suffering in real persons. I have also argued that, other things being equal, we are not morally obligated generally to reduce or prevent the suffering of rational animals or real persons, although we are morally obligated never to cause suffering in them by violating their dignity, namely, to *degrade* them. Furthermore, we are also not morally obligated generally to reduce or prevent the experience of bodily pain, or nociperception, in non-rational human or non-human minded animals. I have also argued that both proto-sentient, simple minded and also sentient, fully minded non-human non-person animals are still primary subjects of moral value and primary targets of our moral concern, and we are therefore obligated, at the very least, to consider them fully, and to take them fully into account in our moral reasoning—which is what I mean by saying that we must treat them as "primary, *serious* targets" of our moral concern.

From all of this, it follows that, other things being equal, it is morally permissible for real persons to treat either human or non-human non-person minded animals in such a way that it foreseeably causes some state of bodily nociperception in them. But at the same time, it is morally impermissible to *torture* them, that is, *treat them with cruelty*, the diametric opposite of *treating them with kindness*. By "torture" and/or "treat with cruelty" I mean the following:

> Conscious subject *X* tortures/treats with cruelty, some proto-sentient or conscious subject *Y* if and only if the choices or actions of *X* are either direct tryings to cause a very high level of bodily nociperception or suffering in *Y*, or, as insofar as *X* is trying to do something else, foreseeably will cause a very high level of bodily nociperception or suffering in *Y*, when it is also really possible to cause a significantly lower level of bodily nociperception or suffering in *Y*.

Therefore, I am saying that it is morally permissible for real persons to treat human or non-human non-person minded animals in such a way that it foreseeably causes *some state* of bodily nociperception in them, although it is morally impermissible either directly to try to cause any highly intense experience of bodily pain in any minded animal, or, insofar as one is trying to do something else, foreseeably will cause a highly intense experience of bodily pain in them when it is also really possible, insofar as that other thing is intended, to cause a significantly less intense experience of bodily pain in them.

Otherwise put, it is a morally impermissible state of affairs whenever someone is trying to cause a highly intense experience of bodily pain in a primary subject of value, or is trying

to cause something else and thereby foreseeably will cause a highly intense experience of bodily pain in that primary subject of value, when he could try to cause a significantly less intense experience of bodily pain in that minded creature. For example, when a veterinarian operates on a dog while benevolently trying to save that dog's life, but also knows that even using anaesthetics, this will inevitably cause a highly intense experience of bodily pain in the dog, then that is morally permissible and not torture or cruelty. But when a medical experimenter investigating the side-effects of a certain drug gives that drug to the very same dog, knowing that it will cause the very same highly intense experience of bodily pain in the dog, but could also choose *either* not to give the drug to the dog at all and achieve the same experimental end by *not* using non-human minded animals, *or else* to give the dog an anaesthetic that would adequately deaden the pain, then that is torture/treating with cruelty, and morally impermissible.

Torture/cruel treatment, thus lies in the moral agent's intention-in-act, *or trying*, not in the consequences of acting on that intention, since in the two hypothetical cases I just described, the dog's very high level of experienced pain is held fixed. In cases like the animal experimentation example, the primary subject of value is being treated either as a mere means, with only the instrumental value of satisfying the torturer's need to hurt other creatures, or as a mere thing, without any sort of moral value, like a piece of garbage or offal, without any moral concern or moral consideration whatsoever, and despite the fact that this primary subject of value possesses at least a constitutively necessary, even if not sufficient, capacity of rational animality or real personhood in common with us. The torturer of non-human non-person minded animals is thereby choosing and acting with "cruelty to animals"; and conversely, s/he who treats non-human non-person minded animals with cruelty is a torturer, and these are equivalently strictly impermissible. So that is what our serious moral concern for non-human non-person minded animals will always morally prohibit, and "kindness to animals" fundamentally consists in heeding this moral prohibition.

I am now finally in a position to raise and answer the following very hard question:

> Is it morally permissible, other things being equal, for us to kill, or cause states of bodily nociperception in, non-human non-person minded animals—including, for example, cephalopods, fish, insects, reptiles, and other proto-sentient or simple minded invertebrates, and also bats, bears, birds, cats, cows, dogs, horses, lions, mice, sheep, and wolves, and other sentient, fully minded non-human non-person animals—for the purposes of, for example, greater human convenience or safety, eating meat, producing other sorts of food, medical experimentation, scientific experimentation more generally, the manufacture of clothing, cosmetics, and furniture, or sport, animal-driven conveyance or transportation (for example, horseback riding, cart-pulling, dog sleds, etc.), or zoos, etc.?

The crucially qualified answer I am offering is Yes, *provided that* this is not torture/cruel treatment. That is: *provided that* this is not either directly to try to cause any highly intense

experience of bodily pain in any minimally minded or conscious animal, or, insofar as one is trying to do something else, foreseeably will cause a highly intense experience of bodily pain in them when it is also possible, insofar as one is trying to do that other thing, to cause a significantly less intense experience of bodily pain in them. For example, these non-torturing/non-cruel-treatment conditions would strongly favor banning or seriously restricting, other things being equal, many current practices of scientific experimention on non-human non-person minded animals, and many current practices of using them in meat production and other sorts of food production, in drug testing, in clothing production, and for display in private zoos, as well as the pointless slaughter of non-human minded animals in traditional sport fishing, big-game hunting, fox-hunting, deer hunting, bird-hunting, and so-on, other things being equal.

On the other hand, however, these conditions would also morally permit, other things being equal, the non-torturing/non-cruel *use* of non-human non-person minded animals in scientific experimentation, meat production, other sorts of food production; the non-torturing/non-cruel *use* of them in animal sports; the non-torturing/non-cruel *use* of them as conveyance or transportation; public zoos; the non-torturing/non-cruel *use* of them as specially-trained companions for people with certain kinds of disabilities; and also the non-torturing/non-cruel *use* of them, via private ownership, as ordinary companions or pets. Correspondingly, these conditions would also morally permit your killing a bee, hornet, or mosquito that is stinging you, or likely to sting you, and also morally permit your killing flies or other insects inside your house, when they are likely to be an annoyance or a health hazard. Nevertheless, they would *still* morally prevent your pulling the wings off flies, or simply killing them (or any other insect, cephalopod, fish, or reptile) slowly and painfully, "for our sport"—like *King Lear*'s cruel gods, or "wanton boys"—other things being equal.

This approach to the treatment of non-human minded animals, in turn, comports very coherently with the widely-held commonsense moral intuition, shared alike by animal liberationists and radical or vegan vegetarians on the one hand, and by non-animal-liberationists and non-radical non-vegetarians who still morally care about our treatment of non-human animals on the other hand, that torturing/cruelly treating non-human non-person minded animals is strictly morally impermissible, no matter what *other* views one may hold about animal ethics. Thus The Concern for All Minded Animals Theory also entails a moral obligation to prevent or reduce cruelty to all minded animals, including of course all non-human non-person minded animals, other things being equal, although it does not *also* entail a moral obligation generally to prevent or reduce harm to them, other things being equal. That latter moral obligation—more specifically, the moral obligation generally to try to prevent or reduce dignity-violating harm, that is, harm which involves someone's being treated as a mere means as a mere thing, without her actual or possible rational consent, and with cruelty—is specially reserved for our respectful treatment of real persons, whether human or non-human. Or in other words:

> Other things being equal, we are morally obligated to try to prevent or reduce the degradation of all real persons; and our serious moral concern for the suffering of all real human persons morally trumps our serious moral concern for the experience of bodily pain in non-human non-person minded animals, assuming roughly comparable levels in the intensity of the experience of emotional or bodily pain, and provided that no minded animal is being tortured/treated with cruelty.

This, in a nutshell, is my Existential Kantian Ethics-based solution to The What-Is-It-Like-To-Be-A-Bat-In-Pain? Problem.

Here, now, is an obvious objection to my solution to The Problem:

> "Suppose that in some or even many act-contexts you are faced with a choice between either (i) preventing or reducing the degradation of some higher-level or Kantian real human person (say, your own beloved partner or child), or (ii) preventing ore reducing the torture/cruel treatment of some non-human non-person minded animal (say, a dog, a bat, an octopus, or even a fly), and both the suffering and the torture/cruelty would involve comparably intense levels of emotional or bodily pain, and you cannot do both? It seems then that your solution leads directly to moral dilemmas."

My reply is that such a situation precisely follows the general pattern of conflicting first-order substantive ceteris paribus objective moral principles that I discussed in chapter 2, and that it is resolved in just the same way—by deploying my Existential Kantian Ethics-based, No-Foolish-Consistency-driven, nonideal Kantian structuralist theory of moral principles, and in particular by deploying The Lesser Evil Principle. In any actual act-context containing conflicting first-order substantive ceteris paribus objective moral principles, I must choose the lesser of the several evils, and more specifically I must choose the course of action in that context which most keeps faith with the Categorical Imperative. Then it is my duty to choose that very course of action in that very context, and there is never a conflict of duties.

Clearly, the ceteris paribus clause in my Existential Kantian Ethics-based, No-Foolish-Consistency-driven, nonideal-Kantian-structuralist-theory-of-moral-principles approach to The What-Is-It-Like-To-Be-A-Bat-In-Pain? Problem, together with The Lesser Evil Principle, effectively rules out apparent counterexamples in which it would seem to be morally impermissible to prevent or reduce the degradation of your own beloved partner or child just because it also involved comparably painfully torturing/cruelly treating a dog, a bat, an octopus, or a fly. The lesser evil in such cases is clearly the prevention of comparably intense levels of degradation in real human person, hence that is your duty in those actual act-contexts, other things being equal, because your partner or child is not only one of your loved ones but also has absolute intrinsic non-denumerable objective value, namely, dignity, whereas a dog, bat, octopus, or fly is neither one of your loved ones nor has dignity.

Nevertheless, it does remain true that, depending on the act-context, these considerations could conceivably morally favor *either* a real human person *or* a non-human non-person minded animal. For example, in Robert Bresson's brilliant 1966 film *Au Hasard Balthazar*, the degradation of a teenage girl (by rape) and the many experiences of bodily pain in the donkey Balthazar are presented in stunningly dramatic parallel. It is not uncommon for viewers of the film to burst into tears when the unfortunate Balthazar dies from ill-treatment, but remain dry-eyed when the equally unfortunate teenage girl commits suicide by rolling herself down a hill, like a log, into a pond.

(It must also be said, parenthetically, that this is a *very* strange way to commit suicide. Yet at the same time, in its own way, it is intensely moving, like so much else in Bresson's films, especially in delayed reaction, as one's capacity for pre-reflectively conscious episodic memory[246] of the suicide scene affectively re-works its essentially non-conceptual content. So it seems to me clear that Bresson is cinematically priming our emotional responsiveness here for Balthazar's death scene a few minutes later.)

In any case, let us then suppose that my choice is between

> (i) preventing or reducing the degradation of the teenage girl, and
> (ii) preventing or reducing the comparably awful torture/cruel treatment of the donkey Balthazar.

Here it is self-evident that preventing or reducing the degradation of the teenage girl most keeps rational faith with the Categorical Imperative in that context, since the Categorical Imperative is specifically designed to protect the dignity of rational minded animals or real persons, especially including all real human persons, and Balthazar is a non-person.

Just to make things even more philosophically difficult, however, now further suppose that, instead, my choice were between

> (i) preventing or reducing the comparably intense suffering of one of the teenage girl's brutally callous tormenters—for example, one of her rapists, now languishing in prison, and seriously depressed, and
> (ii) preventing or reducing the torture/cruel treatment of the donkey Balthazar.

Then it seems much *less* clear that in this context I should choose to prevent or reduce the suffering of the callous tormenter in prison, instead of choosing to prevent or reduce the comparably awful torture/cruel treatment of Balthazar. On the contrary, it seems fairly obvious to me that in this context I should choose to prevent or reduce *Balthazar's* torture/cruel treatment instead. This is because it seems at least plausibly arguable, given the actual content of the film and Bresson's directorial intentions, that the callous rapist's imprisonment would be a morally justified punishment, along with a certain amount of foreseeable likely suffering on his part.

I want to make it emphatically clear, however, that I am *not* saying that that the callous rapist's degradation would be morally justified, if in fact he is a victim of degradation in prison, say, due to his mistreatment by prison guards or other inmates. Nothing would ever morally justify such degradation, even of a very wicked person. All I am saying is that his imprisonment is, foreseeably, likely to lead to some suffering for him.

Furthermore, the callous rapist's punishment—that is, his imprisonment as such, together with some foreseeable likely suffering—would *not* be justified by its being a "just retribution," since in my opinion *retributivism* in the theory of punishment is false.[247] Beyond that, and more radically, *I am an existential Kantian cosmopolitan social anarchist about crime-&-punishment*.[248] Nevertheless, on the (I think, actually *false*) assumption that legal punishment by imprisonment *is ever* morally justified, I think it could be morally justified *only* to the extent that the callous rapist's imprisonment *could be* an effective way of bringing it about that he freely takes complete personal deep moral responsibility for his awful crime, and thereby changes his life for the better.

In any case, the overall coherence and defensibility of my Existential Kantian Ethics-based, No-Foolish-Consistency-driven, nonideal Kantian structuralist theory of moral principles remains intact. This is because the relevant first-order substantive ceteris paribus objective moral principle that is chosen in either case is the lesser of several evils, precisely because it is the one that most keeps rational faith with the Categorical Imperative in that context, and thus there is never a conflict of duties. Correspondingly then, the Existential Kantian Ethics-based solution to The What-Is-It-Like-To-Be-A-Bat-In-Pain? Problem also remains intact.

4.5 KINDNESS TO ALL LIVING BEINGS: ASSOCIATE MEMBERSHIP IN THE REALM OF ENDS

There is one further basic element of The Concern For All Minded Animals view that I still need to explore. According to this view, as I noted in section 3.5 above, under certain conditions—namely, the necessary and sufficient conditions governing the existence and specific character of a normative convention[249]—human or non-human non-persons can be temporarily or permanently treated *as if* they were real human persons falling under the protection of the Categorical Imperative, and thereby gain what I have called "an associate membership in The Realm of Ends." As such, these conventionally-protected creatures are secondary subjects of dignity and secondary targets of respect, and, as extrinsically considered, they receive a temporary or permanent right-to-life, by which, as I said in section 3.4, I mean:

> a subject's unalienable moral demand against others to let her continue being alive, that is, the moral demand not to be impermissibly actively or passively killed by those others, which is not a forfeitable right of any sort, and not a strict right-not-to-be-killed.

Such an associate membership in The Realm of Ends has the following four individually necessary and jointly sufficient conditions.

First, there must be an imaginative extension of the existing ceteris paribus obligation to prevent or reduce dignity-violating harm to real persons (namely, the positive duty to prevent harm), to a pre-selected class of living organisms, whether non-minded, proto-sentient and simple minded, or sentient and fully minded, human or non-human non-persons, where this extension is normally motivated by moral feelings such as compassion, empathy, or sympathy.

Second, there must be a collective rational disposition to provide moral arguments purporting to show that such an extension of specific moral character, aka moral status, is warranted.

Third, there must be an implicit or explicit normative convention between like-minded higher-level or Kantian real persons to confer, defend, and heed this moral status.

Fourth and finally, there must be a generally public, social-institutional recognition of this extension of moral status.

Does this extension of moral protection make good rational sense? Kant says this about *full* membership in The Realm of Ends:

> All rational beings stand under the *law* that each of them is to treat himself and all others *never merely as means* but always also always *at the same time as ends in themselves*. But from this there arises a systematic union of rational beings through common objective laws, that is, a realm (*Reich*), which can be called a realm of ends (admittedly only as ideal) because what these laws have as their purposes is just the relation of these beings to one another as ends and means. A rational being belongs as a *member* to a realm of ends when he gives universal laws in it but is also himself subject to these laws. (*GMM* 4: 433, italics in the original)

Now let us juxtapose *associate membership* in The Realm of Ends and *full membership*. It makes eminently good rational sense that the temporary or permanent possession of a right-to-life by secondary subjects of dignity and secondary targets of respect be in sharp contrast to the possession of dignity by primary subjects of dignity and primary targets of respect—namely, all real persons, including all actualized rational human animals or actualized real human persons, and also all neo-persons. Dignity, with its absolute, non-denumerably infinite, intrinsic, objective value, is an essential property of real persons. But the moral status of associate membership in The Realm of Ends is merely contingent and extrinsic, precisely because it is conventional, although it remains normatively and morally binding

to the extent that the primary subjects of dignity and primary targets of respect are prepared to stand behind it.

A necessary condition for something *X*'s being a secondary subject of dignity and a secondary target of respect is that *X* have a morally valuable *life-of-its-own*, which in turn implies that it must, at the very least, be an individual living organism, or collection of living organisms, for example, a Nature Conservation Zone. So Nero's favorite horse qualifies, and Nero's favorite poisonous snake, and Nero's favorite Venus Fly Trap. But *plastic* (or papier-mâché, or rubber, or metallic, etc.) horses, poisonous snakes, or carnivorous plants do not qualify. Nor do *machines* of any kind—for example, Bob's beloved Bugatti sportscar, to be discussed in the next chapter. This is because being an individual living organism is a constitutively necessary condition of being a subject of dignity and a target of respect.

It is true, as philosophers of art and philosophers of religion have noted, that aesthetic objects, artworks, and sacred objects can also be conventionally and/or intentionally—and for better or worse—assigned an "aura" that is in certain respects quite similar to what I am calling associate membership in the realm of ends.[250] By contrast, the aura of aesthetic objects, artworks, and sacred objects implies at most a *proto-dignity* and not dignity per se.[251] Nevertheless, there are some hybrid cases in which the conventional attribution of associate membership in the realm of ends, and the conventional attribution of the aura of the sacred, coincide, for example, sacred cows in Hindu countries.

Now let us suppose that an associate membership in The Realm of Ends has been actually extended to some human or non-human non-person minded animals or to some other individual living but non-minded organisms. Then, other things being equal, harming those human or non-human minded or non-minded living organisms, for example, by arbitrarily killing them or destroying them, is conventionally morally impermissible. For example, it would be conventionally morally impermissible, other things being equal, arbitrarily to injure or kill your neighbor's cat or dog; arbitrarily to injure or kill minded animals belonging to protected species; arbitrarily to injure or kill insects, bats, or snakes in public zoos; arbitrarily to injure or kill fish or other sea animals in public aquariums; arbitrarily to injure or kill sacred cows in Hindu countries; arbitrarily to injure or kill the bat-fetuses or cat-fetuses of someone's pet bat or pet cat; arbitrarily to cut down or burn the grasses or trees in Nature Conservation Zones, and so-on. And there have, of course, been serious moral debates about extending the same sorts of moral protections to human stem cells or human embryos, on the grounds that they too have morally valuable lives-of-their-own.

Nevertheless, arbitrarily to damage, injure, kill or destroy human or non-human minded or non-minded living organisms in such cases would *not* be a violation of the dignity of that human or non-human non-person minded or non-minded living organism itself, simply because these non-persons do not possess *dignity* per se, but only at most *proto-dignity*.[252] Instead—for example, in the case of arbitrarily injuring or killing your

neighbor's bat, cat, or dog, or arbitrarily cutting down or burning the grasses or trees in a Nature Conservation Zone—it would be at most *indirectly* violating the dignity of those higher-level or Kantian rational human members-in-good-standing of The Realm of Ends who stand behind the moral convention that constitutes this class of associate members of The Realm of Ends, and who jointly confer the status of being a secondary subject of dignity and a secondary target of respect upon those human or non-human minded or non-minded living organisms. Obviously, it would *harm* those organisms. But in this context, provided it is not torture/cruelty, harming those living organisms would directly violate only the conventional moral *office* or moral *role* that is filled or played by those non-persons. Hence it would directly violate no real person's dignity.

As I also noted above, the moral convention whereby secondary dignity and secondary respect, and thereby a temporary or permanent right-to-life, is ascribed to some non-person living organisms derives ultimately from our respect-based moral feelings such as compassion, empathy, or sympathy directed towards all those beings in our world

> that (i) share with us at least one constitutively necessary feature of real personhood—life, but that also (ii) are all non-persons because they lack even the strong potentiality to become real persons.

Associate membership in The Realm of Ends and its corresponding conventional first-order substantive ceteris paribus objective moral principles thus result from coordinated acts of special moral concern and kindness towards minded animals of any species, or towards living organisms of any kind, by real persons like us. And in this way, associate membership in The Realm of Ends provides for what, in effect, is a fairly robust eco-ethical *Noah's Ark Principle* that *could be* endorsed by even the most radical eco-ethicist, for example, Albert Schweizer. For even though Schweizer himself might disagree about its conventionalist metaphysical foundations, *pragmatically speaking*, associate membership in The Realm of Ends and Schweizer's own ethical principles are morally equivalent.

Associate membership in The Realm of Ends is in certain respects similar to Kant's classical "indirect-duty" view, according to which all moral obligations towards non-human non-persons are ultimately obligations towards persons,[253] and not towards non-human non-persons themselves. One important difference, however, is that for Kant's indirect-duty view, the moral obligation to consider and treat non-human non-persons in a certain way is strictly a *duty to oneself*; whereas, according to associate membership in The Realm of Ends, the moral obligation is a duty *to others*. In any case, the two standard objections to Kant's indirect-duty view are

> (i) that it unacceptably implies that we should treat human or non-human non-person minded animals, and in particular all fetuses or infants, either as mere means or as mere things, and

> (ii) that it unacceptably implies that if by some psychological accident torturing/cruelly treating human or non-human non-person minded animals was not bad for me or even improved me (perhaps by releasing aggression), then it would be morally permissible for me to do so.[254]

These objections can easily be rebutted by The Concern For All Minded Animals Theory, when it is added to the notion of associate membership in The Realm of Ends. As we saw above, The Concern For All Minded Animals Theory directly entails that it is morally impermissible to treat any experiencer or subject of moral value in *any* species either as a mere means or as a mere thing, and also that torturing/cruelly treating minded animals of *any* species is as morally impermissible as torturing rational minded animals or real persons, other things being equal. But at the same time, as my Existential Kantian Ethics-based, No-Foolish-Consistency-driven, nonideal-Kantian-structuralist-theory-of-moral-principles solution to The What-Is-It-Like-To-Be-A-Bat-In-Pain? Problem showed, although torturing/cruelly treating minded animals of any species is as morally impermissible as torturing/cruelly treating real persons, other things being equal, it does not follow that minded animals of any species must be *treated equally* with real human persons, that is, treated with equally sufficient respect. On the contrary, other things being equal, the suffering of real human persons morally trumps the experience of bodily pain, aka nociperception, in non-human non-person minded animals, assuming comparable levels of subjectively experienced emotional or bodily pain.

4.7 CONCLUSION

Finally, and even more positively however, one important theoretical advantage of associate membership in The Realm of Ends—insofar as it is a conventional moral mechanism for extending a secondary kind of temporary or permanent moral protection, under the Categorical Imperative, to pre-selected groups of non-human non-person, minded or non-minded living organisms of any kind—is that it thereby avoids the serious problem, for the unconstrained animal liberation theory, of highly implausibly overextending fundamental moral protection to all non-human non-person minded animals *in the wild*, and particularly *in natural predation situations*. Indeed, the highly plausible explicit or implicit moral belief *that there is a basic asymmetry between the fundamental moral protections applying to real human persons on the one hand, and the moral protections extended to non-human non-person minded animals in the wild and in natural predation situations on the other*, is shared by *all* parties to the debate about the morality of our treatment of non-human minded animals, including the most radical eco-ethicists, for example, Schweizer. Schweizer would have tried to protect any real human person who was being attacked by another, and would also have tried to prevent any such attacks, but

he did not try to stop, nor did he urge us to prevent, natural predation among non-human, non-person animals. Scweizer's moral belief entails that we do *not* have an obligation, other things being equal, *to prevent or reduce the experience of bodily pain in the wild and in natural predation situations, or to prevent the arbitrary killing of such animals by one another in such situations*, whereas we *do* have an obligation (of some sort), other things being equal, *to prevent or reduce the degradation of real human persons, and to prevent the arbitary killing of real human persons by one another, in all situations*. In turn, this universally shared moral belief clearly supports The Concern For All Minded Animals Theory. And it also clearly supports the important sub-thesis of The Theory, which says that the commonplace inference from the fact of bodily nociperception in animals to their suffering—namely, The Bentham-Singer Fallacy—really *is* a fallacy.

For not even the most radical eco-ethicist, not even Schweizer, would be rationally prepared to say that when, in the ordinary course of natural predation, a mountain lion kills and eats a deer, then that deer is thereby suffering in the precise and morally weighty sense of that term, *so that, other things being equal, we are morally obligated to stop the mountain lion if we can, or to prevent that natural predation from occurring, including using lethal force if necessary*, even though, obviously, that deer still *is* experiencing intense bodily pain.

Chapter 5

TROLLEYS, BRIDGES, HUMAN MISSILES, AND PONDS: THE MORALITY OF SAVING LIVES

> As utilitarians tirelessly and rightly point out, very rarely should ordinary agents (as opposed to trolley operators) think they can produce large net benefits only by harming innocent others. In contrast, given the effectiveness of UNICEF, OXFAM, and similar agencies which aim to prevent death and minimize pointless suffering, the opportunities most ordinary agents (or at least most people in relatively rich countries) have to make sacrifices in order to rescue others are ubiquitous. Thus the question of how much we are required to sacrifice has everyday relevance. Because so much is at stake—large numbers of innocent lives—and so many people regularly have opportunities to help, the central question [here] is the most important one in contemporary normative ethics.[255]

> When Bob first grasped the dilemma that faced him as he stood by that railway switch, he must have thought how extraordinarily unlucky he was to be placed in a situation in which he must choose between the life of an innocent child and the sacrifice of most of his savings. But he was not unlucky at all. We are all in that situation.[256]

5.1 INTRODUCTION

What I will call *The Problem of Saving Lives* is this:

> How much am I morally permitted or obligated to do in order to save the lives of some mortally threatened real human persons, whether others or myself?

In this chapter, I shall argue for a three-part solution to that problem that runs as follows.

First, in view of the fact that those mortally threatened real human persons are ends-in-themselves who have absolute, non-denumerably infinite, intrinsic, objective moral value, aka *dignity*, together with the five further facts that

(i.1) I am obligated, other things being equal, not to harm real human persons by violating their dignity,
(i.2) I am obligated, other things being equal, to prevent or reduce dignity-violating harms to real human persons,
(i.3) I am obligated, other things being equal, to prevent or reduce the degradation of real human persons,
(i.4) I am obligated, other things being equal, to promote the happiness of real human persons, and
(i.5) I am obligated, other things being equal, to develop my abilities and perfect myself,
then it follows that
(i.6) other things being equal, I am permitted or obligated to do *quite a lot* in order to save the lives of mortally threatened real human persons, whether others or myself.

Second, this "quite a lot" specifically includes

(ii.1) being permitted to kill a few real human persons in order to save a significantly greater number of others, provided that
(ii.1a) no innocent bystander is being treated as a mere thing, even though, in some very special cases in which "other things are not equal," she *is* morally permissibly treated as a mere means, and
(ii.1b) no innocent bystander is being treated without her actual or possible rational consent, and also
(ii.2) being permitted to kill another real human person in self-defense, even if he is only an innocent attacker, provided that
(ii.2a) killing is the only way I can protect myself from being mortally threatened in that context, and that
(ii.2b) only minimal lethal force is used by me, and also
(ii.3) being personally morally deeply responsible[257] for sacrificing something that is of moral significance to me, not to mention also being personally morally deeply responsible for sacrificing things that are of some moral value to me but not of any moral significance to me, in order to save the lives of some other real human persons, provided that
(ii.3a) I am the closest one in space and time, in that context, among all the close ones,[258] to the mortally endangered real human persons,
(ii.3b) I am the only one who can save these mortally threatened real human persons in that context, and
(ii.3c) I am not morally required to repeat this act of sacrifice to the point at which it undermines my obligatory life-project, other things being equal, of developing my abilities and perfecting myself, and
(ii.4) being obligated to sacrifice at least sometimes something that is of some moral value to me, but of no moral significance to me, in order to save the lives of some other real human persons, even if
(ii.4a) I am not the closest one in space and time to the mortally endangered real human persons in that context, and

(ii.4b) I am not the only one who can save these mortally threatened real human persons in those contexts, provided that
(ii.4c) that I am not morally required to repeat this act of sacrifice to the point at which it undermines my obligatory life-project, other things being equal, of developing my abilities and perfecting myself.

Third, precisely how much I am permitted or obligated to do in order to save these mortally threatened real human persons will depend crucially on certain ineluctably contingent contextual features of the relevant threat-situations, which may or may not make it my *personal moral deep* responsibility[259] to save them in that situation, which may or may not place some innocent real human person in a position to be morally permissibly sacrificed in that situation, and which may or may not place some innocent real human person in a position to be morally immune to sacrifice in that situation. More specifically, this third part of the answer invokes something I call *The Morality De Re Thesis*, which says:

The normative contents of first-order substantive ceteris paribus objective moral principles are partially determined by ineluctably contingent spatial, temporal, and causal contextual factors, *as well as* emotional and social contextual factors such as love-bonds, friendship-bonds, and family-bonds, in the concrete situations in which choice or action occurs according to these principles, which

(i) sometimes make innocent higher-level or Kantian real human persons, individually, morally deeply responsible for saving others, so that saving others is just "up to them" in those concrete situations, even at a surprisingly high moral cost to themselves,
(ii) sometimes place innocent real human persons in a position to be morally permissibly either killed in self-defense or sacrificed for the sake of others, so that they are just "innocent casualties" of those situations, and
(iii) sometimes place real human persons in a position to be morally immune to sacrifice, so that they are just "innocent bystanders" *alongside* those situations.[260]

In other words, the morality of saving lives is partially determined by *moral luck*.[261]

Corresponding to the Morality De Re Thesis, I will call this three-part solution to The Problem of Saving Lives, *The Morality De Re Solution*. If The Morality De Re Solution is cogent, then it has a large theoretical pay-off, since it thereby also offers a unified solution to three outstanding problems in contemporary normative and applied ethics:

(i) *The Trolley Problem*, as originally developed by Philippa Foot and Judith Jarvis Thomson,
(ii) *The Self-Defense Problem*, as originally formulated by Thomson, and
(iii) *The Famine Relief Problem*, as originally developed by Peter Singer and Peter Unger.

And in the course of providing a joint solution to that triad of problems, The Morality De Re Thesis *also* yields an adequate solution to a hard problem faced by anyone who is thinking seriously about any of these topics,

(iv) The (Im)Partiality Problem:

> How can I be morally justified in saving the lives of those real human persons who are tied to me by special emotional and social bonds, hence acting on the basis of my partiality for these people, over those real human persons who are equally mortally threatened but are also strangers to me, and yet still satisfy the moral demands of impartiality?[262]

That fourfold solution, in turn, provides significant further justification for Existential Kantian Ethics, precisely because it thereby directly existentially, practically, and theoretically binds together what had otherwise seemed to be four relatively detached and distinct, hence recherché and merely "scholastic" and "casuistic," although in fact *profoundly problematic*, bits of mesh in The Web of Mortality.

5.2 Runaway Trolleys and Bridges: The Trolley Problem

As Kant correctly pointed out, since

(i) all real human persons have dignity, and since
(ii) all real human persons naturally desire happiness, and since
(iii) all real human persons belong to The Realm of Ends and morally owe each other equal consideration, then it follows that
(iv) I have a duty to promote the happiness of all other real human persons:

> Concerning … duty to others, the natural end that all men have is their own happiness. Now humanity might indeed subsist if nobody contributed anything to the happiness of others, provided he did not intentionally impair their happiness. But this, after all, would harmonize only negatively and not positively with *humanity as an end in itself*, if everyone does not also strive, as much as he can, to further the ends of others. For the ends of any subject who is an end in himself must as far as possible be my ends also, if that conception of an end in itself is to have its *full* effect in me. (*GMM* 4: 430, italics in the original)
> When it comes to my promoting happiness as an end that is also a duty, thus must therefore be the happiness of *other* men, *whose* (permitted) *end I thus make my own end as well*. (*MM* 6: 388, italics in the original)

But this seemingly unexceptionable principle, even when it is explicitly taken by Existential Kantian Ethics to be a first-order substantive ceteris paribus objective moral principle, leads to a very hard moral-philosophical problem. Correspondingly, in "Abortion

and the Doctrine of the Double Effect,"[263] Foot spelled out what is now commonly known as The Trolley Problem:

> If it is judged morally *permissible* for the driver of a runaway trolley which is heading straight towards five innocent people to turn his trolley onto a spur occupied by one innocent person, thereby killing one in order to save five (aka *Trolley Driver*), then why is it judged *impermissible* for a surgeon to kill one innocent patient and distribute his organs to five other dying patients in order to save those five people (aka *Transplant*)?

Otherwise put: What is the morally relevant difference between *Trolley Driver* and *Transplant*? If, to almost everyone who rationally considers these thought-experiments,[264] it seems *sometimes* morally permissible to kill a few innocent real human persons, and therefore create a few innocent casualties, in order to save significantly more innocent real human persons from dying—thereby choosing and acting not only in accordance with the first-order substantive ceteris paribus objective Kantian moral principle to promote the happiness of other real persons, which is also in at least superficial conformity with altruistic act-utilitarian principles—then why is it not *always* morally permissible? This is what I will call *The Trolley Problem of Saving Lives*.

Now Foot's own proposed solution to The Trolley Problem of Saving Lives is to use the killing vs. letting die distinction,[265] and claim that

> (i) killing one is worse than letting five die, and
> (ii) killing five is worse than killing one.

Put in terms of the classical negative duties vs. positive duties distinction, Foot is saying that it is worse to violate a negative duty not to harm one, than it is to violate a positive duty to save five, and also that it is worse to violate a negative duty not to harm five, than it is to violate a negative duty not to harm one. The trolley driver in *Trolley Driver* will either kill five or kill one, so he must kill the one. By contrast, the surgeon in *Transplant* will either kill one or let five die, so he must let five die.

But as Thomson very effectively shows in "Killing, Letting Die, and the Trolley Problem" and its sequel "The Trolley Problem,"[266] this cannot be correct. That is because of the following variant on *Trolley Driver*, which seems clearly morally permissible to almost everyone who rationally considers it:

> You are standing beside a switch at the spur, which you know how to operate, and you see the runaway trolley without a driver, so you decide to turn the trolley onto the spur and thus onto the one, so that you kill him, thereby saving the five (aka *Bystander at the Switch*).

Your choice here is between killing one and letting five die, hence it is false that it is always worse to kill one than to let five die. So, given the clear permissibility of *Bystander at the Switch*, the killing vs. letting die distinction does not solve the Trolley Problem.

Now at this point an orthodox Kantian might appeal to the Categorical Imperative's Formula of Humanity as an End-in-Itself, namely,

> 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 (*GMM* 4: 429),

and say that the moral difference between *Trolley Driver* and *Transplant* is that whereas the surgeon in *Transplant* uses the one "merely as a means" to saving five, neither the trolley driver in *Trolley Driver* nor the bystander in *Bystander at the Switch* uses the one "merely as a means" to saving the five. If the one by some miracle disappears or is otherwise removed from the track (say, by a swooping air-sea rescue helicopter) before the trolley reaches him, then the intentions of the trolley driver's act and the bystander's act are satisfied just the same.

But unfortunately for orthodox Kantians, this suggestion is refuted by Thomson's ingenious "loop variant"on *Bystander at the Switch*.[267] This variant extends the spur and loops it around back onto the five, and thus *causally requires* the death of the one—presumably, the automatic brakes of the trolley are triggered as it runs over the one—as a means to saving the five, hence it treats the one *as a mere means* to saving the five. Now to almost everyone who rationally considers the loop variant, it seems as morally permissible as *Bystander at the Switch*. Thomson considers the idea that what accounts for the morally relevant difference between *Trolley Driver* and *Transplant* is the fact that some "right in the cluster of rights one has in having a right-to-life"[268] of the one is violated in *Transplant*, but not violated in *Trolley Driver*. She calls this sort of right a "stringent right," because it is a non-interference right or liberty right not to be harmed. But this does not seem to be sufficient, since both in *Trolley Driver* and *Bystander at the Switch* it appears that some "right in the cluster of rights one has in having a right to life" of the one is in fact violated. Thomson then draws our attention to another salient difference between *Trolley Driver* and *Bystander at the Switch* on the one hand, and *Transplant* on the other:

> In *Trolley Driver* and *Bystander at the Switch* an existing threat is deflected from five onto one, whereas in *Transplant* a new threat is introduced and imposed on the one.

Nevertheless, at this point in Thomson's argument, it is somewhat unclear just *why* the distinction between deflecting old threats and introducing new threats makes a genuine *moral* difference.

One distorting feature of both the original *Trolley Driver* and *Transplant* cases is that trolley drivers and surgeons may, by virtue of their social roles, have positive duties to

provide certain goods and services for other people. *Bystander at the Switch*, by contrast, is morally analogous to *Trolley Driver* but does not include this distorting feature. Therefore Thomson introduces a corresponding non-distorting analogue of *Transplant*, which she calls *Fat Man*:

> You are standing on a footbridge over the trolley track. Beside you is a really fat man. You see the runaway trolley below you heading towards the five, and realize that if you push the fat man down onto the tracks he will stop the trolley. So you decide to push the fat man off the bridge and save the five. [269]

Almost everyone who rationally considers these cases thinks that your choice or act in Fat Man is morally impermissible. So now the non-distorting version of The Trolley Problem is this:

> What is the morally relevant difference between *Bystander at the Switch* and *Fat Man*?

By way of a proposed solution Thomson offers the following principle, which I will call *Thomson's Trolley Principle*, by combining her thought about rights violations with her thought about deflecting threats:

> It is morally permissible to kill one in order to save five if and only if we do so by deflecting an existing threat in such a way that the act of deflection itself violates no stringent right of the one.

Otherwise put, Thomson is saying that it is permissible to kill a few innocent people in order to save significantly more innocent people, as long as we introduce no new threats *and also* do not violate any stringent rights in the means we use to get the existing threat onto the one.

I think that Thomson's Trolley Principle clearly fails. This can be seen in a case I will call, for lack of a more elegant label, *the shoving-the-small-person-aside variant* on *Bystander at the Switch*.[270] In this variant, everything is the same as *Bystander at the Switch*, except that there is now one innocent small human person standing in front of the switch, and no one on the spur, and you cannot get to the switch except by shoving her aside. Sadly, the small person then falls right in front of the trolley and is killed. You do not *try* to push her in front of the trolley, and in fact you shove her with only the minimum amount of force necessary to clear her out of the way. Nevertheless, you do foresee that because she is so very small, it is almost inevitable that she will fall that way and be killed, yet you go ahead and shove her aside anyway, and she is killed. So you are using and treating, and in fact killing, the small person "merely as a means" to saving the five. It seems very clear that if *Bystander at the Switch* and the loop variant are both morally permissible, then so is the shoving-the-small-person-aside variant. The small real human

person, sadly, is just an innocent casualty of that mortal threat situation. But the shoving-the-small-person-aside variant violates a stringent right of the one in the act of deflecting that threat, namely, her right-to-life, not to mention treating her merely as a means. So neither Thomson's Trolley Principle nor orthodox Kantian ethics correctly isolates the morally relevant difference between *Bystander at the Switch* and *Fat Man*. But if Thomson's Trolley Principle and orthodox Kantian ethics both fail, then what is the correct solution to The Trolley Problem of Saving Lives? In my view, the correct solution is The Morality De Re Solution.

According to The Morality De Re Solution to The Trolley Problem of Saving Lives, it is crucial to recognize that the dignity of real human persons does *not* entail that it is *absolutely always* morally impermissible to treat people as a mere means, as both the loop variant and the shoving-the-small-person-aside variant on *Bystander at the Switch* clearly show.[271] This is for two reasons.

First, even though the person who is sacrificed is being *treated* as a mere means, nevertheless at the same time it is strictly "nothing personal," in the sense that it is not in any way required that the actual unique life of *this* or *that* person be destroyed in order to save five other innocents. If on the contrary it were "something personal," and the small person were to be *specially selected for sacrifice*, as in *Fat Man*, and as it were pulled out from behind the Rawlsian veil of ignorance[272] and made to take a fatal hit, then that would be treating her like a mere thing, and thereby harming her by violating her dignity as a real human person.

Again, and now put in terms of the well-known *de re vs. de dicto* distinction in philosophical logic,[273] it is not specifically required of *her*, the very person that she actually is, in this context, that *she and she alone* become an innocent casualty in order to save five (de re). Rather it is only required that *someone or another*, *who just happens to be her* in this context, become an innocent casualty in order to save five (de dicto). Let us call this *The Nothing-Personal Criterion*, and it specifically captures *the de dicto standpoint of the harming agent, or sacrificer*, as he looks into causally-accessible nearby possible act-worlds in order to find some way of saving the five.

Second, the person who is sacrificed can give her actual or possible rational consent to being treated in this way. Behind the Rawlsian veil of ignorance, would you not rationally consent to any moral world in which, under very special crisis-conditions in which "other things are not equal," one innocent person can be used as a mere means and killed in order to save five others, even if there were a small chance of your being accidentally the one who is on the spot—if the alternative is a moral world in which this either never happens, or else only ever sometimes happens? It seems clear that you should answer this affirmatively. Let us call this *The Rational Consent Criterion*, and it specifically captures *the de re standpoint of the harmed agent, or sacrificial victim*, as she looks towards the oncoming threat and asks herself whether in some causally-accessible nearby possible act-world she would be willing to lay down her life in order to save the five.

Taken together, The Nothing Personal Criterion and The Rational Consent Criterion tell us when innocents may be permissibly harmed or even killed in order to save mortally endangered others, and when this is impermissible.

The Morality De Re Solution to The Trolley Problem of Saving Lives says that the absolute, non-denumerably infinite, intrinsic, objective value, aka dignity, of real human persons entails that it is *absolutely always* morally impermissible to treat people *either as mere things or without their actual or possible rational consent*, even if, in some very special contexts in which "other things are not equal," it *is* morally permissible to treat them as a mere means. This entailment carries over directly into the semantic content of The Nothing Personal Criterion and The Rational Consent Criterion alike. Both of these criteria are violated in *Fat Man*. By virtue of some ineluctably contingent contextual differences in proximity, distance, and causation, the Fat Man is originally placed in a position to be an innocent bystander alongside an ongoing mortal threat situation, and is morally immune to sacrifice with respect to that situation. But he has nevertheless been forced into that very situation and is being treated "as a mere trolley-stopping thing," and not as an end-in-himself, and thereby he is also being treated without his actual or possible rational consent. So the trolley-stopping has become something degradingly personal for him: in that context, it *has to be* the Fat Man who does it. Otherwise put, in that context, he is being treated as *nothing but a sufficient causal trigger for the trolley's brakes*, so that the five can be saved, and the greater good promoted. So the Fat Man can rationally fully expect never to receive a sufficiently justified answer to the question: "*Why* does it have to be *me*?," and therefore he will refuse to give his actual or possible rational consent to being sacrificed. Hence it is morally impermissible to kill the Fat Man by pushing him off the bridge and down in front of the trolley.

By sharp contrast, the one in *Bystander at the Switch*, the loop variant, and the shoving-the-small-person variant are all placed in positions to become innocent casualties of those ongoing threat-situations, for whom The Nothing Personal Criterion and The Rational Consent Criterion are both satisfied, although this still happens by virtue of some ineluctably contingent contextual differences in proximity, distance, temporality, and causation. None of them can ask "*Why* does it have to be *me*?" and rationally fully expect never to have a sufficiently justified answer. This is simply because (for example, in the sideways-shoving variant) it did not *have to be* her. It was nothing personal at all. She just happened to be in the wrong place at the wrong time. So, sadly, it was just *very bad moral luck*.

What is essential for *Bystander at the Switch*, the loop variant, and the shoving-the-small-person variant is that it is *no*t the intention of the harming agent or sacrificer to kill the one as "something personal." It is *not* the intention of the harming agent to treat the one "as a mere trolley-stopping thing," or as a mere causal trigger for the trolley's brakes, in order to bring about the greater good of saving the five. If, counterfactually, there had been any other way of saving the five without sacrificing the one's life, then the harming agent

would have used that other causal means to the greater good. Hence the one can morally permissibly be sacrificed.

This set of moral-luck based, Categorical Imperative-based, and counterfactual moral properties characteristic of The Morality De Re Solution also constitutes what Frances Kamm (fairly opaquely) calls "the noncausal flip side" of the sacrificial causal means, in the actual sequence, of bringing about the greater good of saving the five. This notion of a noncausal flip side, in turn, is an essential feature of her proposed "Principle of Permissible Harms" (PPH):

> The basic idea of the PPH is that an act is permissible if (i) a greater good or (ii) a means that has a greater good as its non-causal flip side causes a lesser evil. However it is not permissible for an act (iii) to require a lesser evil (or someone's involvement leading to a lesser evil) as a means to a greater good or (iv) to directly cause a lesser evil as a side effect when it has a greater good as a mere causal effect unmediated by (ii). By "noncausal flip side" is meant that the description of the occurrence of the means to the good (i.e., the turning of the trolley) in a context in which there are no other threats to the five is also a description of the five being saved and hence a description of the occurrence of the greater good…. The PPH should [also] be revised to allow that a structural equivalent of the greater good or a means that has it as a noncausal flip side may produce a lesser evil, even when the lesser evil is necessary to sustain the greater good (by defusing new problems that arise from possible remedies for the original threat).[274]

Kamm's PPH-based solution and The Morality De Re Solution to The Trolley Problem of Saving Others are, I think, extensionally equivalent across actual and possible cases and therefore minimally consistent with one another. There are two basic non-extensional, aka *intensional*, differences between the two solutions, however, both of which strongly favor The Morality De Re solution.

First, The Morality De Re Solution solution is superior in a justificatory or reasons-giving sense. This is because it explicitly makes a categorically normative appeal to The Nothing Personal Criterion and The Rational Consent Criterion, both of which, in turn, directly invoke the Categorical Imperative and the No-Foolish-Consistency-driven, nonideal Kantian structuralist theory of moral principles that I presented in chapter 2. By contrast, although Kamm's PPH-based solution is officially "non-consequentialist," this apparently is by stipulation only.[275]

Second, The Morality De Re Solution is also explanatorily superior. This is because it is explicitly grounded in a robust background metaphysics of free agency and real human personhood.[276] By contrast, Kamm's stipulatively non-consequentialist PPH-based solution rests entirely on the dangerously thin and un-reinforced scaffolding of common sense moral intuitions and reflective equilibrium alone,[277] unsupported by any deeper or independent rationale.

Obviously, much turns here on the morally fundamental idea, captured in The Formula of Humanity as an End-in-Itself formulation of the Categorical Imperative, that it is absolutely always impermissible to treat anyone *as a mere thing*, because this violates that real person's dignity. So more precisely then, what does it mean to treat someone "as a mere thing"?

At the very least, it seems that to treat someone else *as a thing* is to regard or treat the other person *as if he or she had no dignity*. It is therefore to regard or treat the other person as if whatever moral value he or she had was at best

> either (i) relatively intrinsic, such that his or her value as an end is *solely* the result of my desiring the existence of that person or some property of that person in a self-interested or aesthetically disinterested way,
> or else (ii) merely extrinsic, such that his or her value is *merely instrumental* to some further relatively intrinsic self-interested or aesthetically disinterested moral value.

In turn, there are at least two different ways of having merely extrinsic or instrumental moral value.

The first way involves the idea that *X*'s being treated as a mere means, that is, as a mere instrument or *tool*, also requires, or at least does not inherently rule out, the continued existence and functionality of *X*. I will call this *re-usable* merely extrinsic or instrumental moral value.

And the second way involves the idea that *X*'s being treated as a mere means not only does not require the continued existence and functionality of *X* but in fact also strictly rules out this continued existence and functionality, by entailing the consumption or destruction of *X*. I will call this *disposable* merely extrinsic or instrumental moral value.

Now the very special cases in which "other things are not equal" and someone can be morally permissibly treated as a mere means, are *all* cases in which the person being so treated is regarded or treated as having *re-usable* merely extrinsic moral value, but not as having *disposable* merely extrinsic moral value. By sharp contrast, mere things have no intrinsic moral value whatsoever, whether absolute or relative. Furthermore, mere things do not have a *re-usable* merely extrinsic value. Therefore, insofar as they have merely extrinsic moral value, mere things have only *disposable* merely extrinsic moral value. Mere things are, in this respect, nothing but fodder or fuel for people's lives insofar as they are governed by purely instrumental norms and purely instrumental practical reasoning.

Moreover, some mere things do not have any *positive* extrinsic moral value at all but rather only *negative* extrinsic moral value. This can happen in several different ways. Mere things may be simply obstinately useless and need to be washed or swept away, like dirt or dust. Or mere things may be disgusting or noxious, and need to be exterminated, discarded, or flushed away, like pestilence, garbage, or offal.

Then to treat someone as a mere thing would be to treat that real person as if she were nothing but fodder, fuel, dirt, dust, pestilence, garbage, or offal. It is misnamed "dehumanization" because it is in fact *de-personalization*. More specifically it is *real-human-person-degrading*, and if its victims do survive, it also produces in them the very worst kind of suffering, the suffering of degradation (see section 4.4 above). In any case, this sort of treatment of real human persons is pretty much the most horrible thing in the world. It is how the Nazis actually and systematically treated millions of people, and horrifyingly, of course, they are not the *only* ones to have done so since the mid-1930s, even on comparable scales of magnitude. To take just one example, it is clear that Americans and Japanese certainly regarded each other, and also more or less systematically treated each other, in this very way during the brutal Pacific War from 1941-1945.[278] And the list of such abuses since that time goes on and on and on.[279]

In this connection, I should also note that it is perfectly consistent with the near-satanically evil mindset of those who treat other people as mere things in either the obstinately useless or the disgusting, noxious senses, that the victims of this treatment be used up, washed or swept away, exterminated, discarded, or flushed away by so-called "humane methods" involving anesthesia, highly efficient pest-control techniques, extreme cleanliness, or "best practices"—as in the hideously sanctimonious sign over the entrance to Auschwitz, and other death-camps, *Arbeit macht frei*, "work makes you free," or in the equally hideous term, "ethnic cleansing," used by near-satanic fellow travellers of the Nazis well after 1945, and now well into the 21st century.

The other idea that needs further explication here is the notion that ineluctably contingent contextual differences in proximity, distance, and causation, as well in egocentrically-centered emotional and social relations, can partially determine the content of first-order substantive ceteris paribus objective moral principles in the contexts in which they are chosen or acted upon. These contextual differences in proximity, distance, temporality, and causation, as well in egocentrically-centered emotional and social relations, are not, in and of themselves, morally relevant differences, no matter how naturally or personally important they might otherwise be. But they do *pick out and trigger morally relevant differences*, that is, they pick out contextual differences that trigger the specific application of first order substantive ceteris paribus objective moral principles and the concrete determination of moral duties.

This is closely analogous to the phenomenon of "essential indexicality" in the philosophy of language and in the theory of mental content,[280] which I have also explicated elsewhere in terms of essentially non-conceptual content.[281] The fact of essential indexicality is in play whenever the semantic content of a term or judgment displays inherent, systematic, and non-reducible context-sensitivity. Thus, for example, I can be the referent of "I" in some cases, and the referent of "he" or even "that" in others, and it is only contextual differences in spatial or temporal position, varying from case to case, that trigger semantic reference.

There are many interesting truths of the logic and semantics of essential indexicality. For example, necessarily, I am the only person in the world who can say "I" and mean *me*; and necessarily, each distinct speaker who uses "I" refers to a different individual person, namely that very speaker. Moreover, even though the terms "I," "he," and "that" systematically shift reference depending on the context, they mean different things at the level of what is known as the *character*, or general semantic function, of those expressions.[282] Even though, necessarily, the contextual referent of "I" is me, the character of "I" is, roughly, *whoever is here and now using this token of the word 'I'*. Correspondingly, the character of "he" is, roughly, *whoever is the male minded animal indicated by the speaker here and now*, And the character of "that" is, roughly, *whatever is now over there in the place indicated by the speaker*. So too with proximity, distance, temporality, and causation, as well as egocentrically-centered emotional and social relations: they vary by context, but they are necessary features of every moral situation. In turn, it follows directly from these points that proximity, distance, temporality, and causation, as well as egocentrically-centered emotional and social relations, are all systematically contextually morally relevant and cannot be explained away. So let us call this *moral essential indexicality*. Moral essential indexicality entails the necessary presence of essentially non-conceptual content in at least some and perhaps all objective moral principles and moral judgments.

Moral essential indexicality has one other crucial feature, namely, what is called *egocentric centering*. This means that the moral relevance of spatial, temporal, and causal factors, as well as emotional and social relations, via what Maiese and I call "affective framing,"[283] in a given context or situation, always depend *on establishing a subjectively experienced center or origin-point in that context/situation*. Relatively to that subjectively-experienced center or origin-point (that is, a spatiotemporal and causal point of view, or affective frame), the contextually/siutationally determined factors of proximity, distance, time, and causation, as well as emotional and social relations (*I* vs. *thou*, *Us* vs. *Them*, and so-on), can then all be suitably *morally calibrated*. So actual and possible moral contexts or situations are an example of what are called *centered possible worlds*. But the moral calibration can vary from context/situation to context/situation. What counts as morally relevant proximity in one context/situation (say, being two feet away from someone who is standing in front of a track-switching device, or standing next to a loved one, friend, or family member, as opposed to a stranger), may or may not count as morally relevant proximity in another. Each new context/situation needs to be morally *re-calibrated*.

Let's come back now to The Trolley Problem of Saving Lives, treated according to The Morality De Re Solution. The crucial moral difference between *Bystander at the Switch* and *Fat Man* is essentially indexically determined by the sacrificed/trolley-killed one's brute spatial distance or proximity, temporal overlap, and causal relatedness to the ongoing *mortal-threat-situation*, together with the sacrificer/switch-throwing bystander's not treating anyone either as a mere thing or without his actual or possible rational consent,

by forcing him or her from an otherwise "causally buffered," or relatively causally inaccessible, spacetime position into the relvant ongoing mortal-threat-situation, thereby killing him or her. This leads to the following principle:

> Other things being equal, it is morally permissible to kill one innocent real person in order to save five other innocent real persons if the one is already a participant in an ongoing mortal-threat-situation, provided that
>
> (i) no innocent bystander is being treated as a mere thing by being forced into that mortal-threat-situation from an otherwise causally buffered spacetime position, and
> (ii) no innocent bystander is being treated without her actual or possible rational consent by being forced into that mortal-threat-situation from an otherwise causally buffered spacetime position.

I will call this *The Specific Morality De Re Trolley Principle*. The Specific Morality De Re Trolley Principle adequately explains the morally relevant difference in all the pairs of specific 1-person- vs.-5-person cases covered by Thomson, including the cases that count against her theory—that is, including the shoving-the-small-person-aside variant—and other similar cases. More generally however, I am saying:

> Other things being equal, it is morally permissible to kill a few innocent real persons in order to save significantly more innocent real persons if the few are already participants in an ongoing mortal threat situation, provided that
>
> (i) no innocent bystander is being treated as a mere thing by being forced into that mortal threat situation from an otherwise causally buffered spacetime position, and
> (ii) no innocent bystander is being treated without her actual or possible rational consent by being forced into that mortal threat situation from an otherwise causally buffered spacetime position.

I will call this *The Generalized Morality De Re Trolley Principle*. This principle is extensionally equivalent with Kamm's Principle of Permissible Harms, but it is also intensionally *non*-equivalent, and, as we saw above, it is arguably *significantly more defensible* than Kamm's Principle of Permissible Harms in both the justificatory and also explanatory senses—not to mention, frankly, its also being significantly easier to understand than Kamm's fairly-opaquely-formulated Principle.

One possible objection to both of The Morality De Re Trolley principles is that the innocent casualties in the loop variant and in the shoving-the-small-person-aside variant do not differ in any morally relevant way from the innocent bystander who is immune to sacrifice in *Fat Man*. After all, the objector would say, in each case the one is treated "as a mere means" to saving the five, whenever s/he is sacrificed in order to save the five.

But I would reply that the Fat Man's also being treated as a mere thing and without his actual or possible rational consent, and correspondingly his also being treated as disposable, and like fodder or fuel for an all-purpose utility-maximizing machine—even if he is not treated in this context as obstinately useless, or as disgusting and noxious, like dirt, dust, pestilence, garbage, or offal—does not consist in his trolley-stopping role alone. It is his trolley-stopping role, *together with* the further fact that by virtue of ineluctably contingent factors of proximity, distance, temporality, and causation, he is positioned by the world to be an otherwise causally buffered innocent bystander, minding his own business, who is then forced into being sacrificed and treated as "disposable dry goods," that jointly entail his being impermissibly treated as a mere thing and without his actual or possible rational consent. And in this way, killing the Fat Man fails to satisfy either The Nothing Personal Criterion or The Rational Consent Criterion. Or in other words, it is the fact that the Fat Man is forcibly moved by me from one kind of essentially indexically determined moral status (namely, that of an innocent bystander, with immunity to sacrifice) to another kind (namely, that of an innocent casualty, without immunity to sacrifice), that entails his being impermissibly treated as a mere thing and without his actual or possible rational consent.

Another possible objection is closely related to this idea. It says that in treating the one in *Fat Man* as a real human person with dignity, we must respect the autonomous choices of the one. We might then plausibly think that it would be morally permissible and even highly morally praiseworthy, although supererogatory, for the Fat Man to choose to throw himself down onto the tracks in order to save the five. But if so, then how could the Fat Man fail to be able to give his actual or possible rational consent to my pushing him down onto the tracks?

My reply to this objection would be that it is one thing for the Fat Man to consider giving his actual or possible rational consent to *his jumping down* onto the tracks in order to stop the trolley, and quite another thing altogether to consider giving his actual or possible rational consent to *his being pushed down* onto the tracks in order to stop the trolley. The former he indeed *can* rationally consent to, as a morally great and heroic, although non-obligatory, autonomous act of self-sacrifice (see section 6.7 below)—but the latter he *cannot* rationally consent to. More precisely, the Fat Man cannot rationally consent to his being treated as a mere thing, whether by himself or by others, simply because it is absolutely impermissible to treat *anyone* as a mere thing, even if the causal and moral value consequences of his jumping down onto the tracks and his being pushed down onto the tracks are exactly the same. The Fat Man's being pushed down onto the tracks directly violates the Formula of Humanity as an End-in-Itself in its self-directed/reflexive application, and this remains the case "whatever the consequences."

5.3 HUMAN MISSILES AND MORE BRIDGES: THE SELF-DEFENSE PROBLEM

The most obvious and widely-supported counterexample to the commonsense claim that killing others is always morally impermissible, is the case of permissible killing in self-defense when the attacker is trying to kill you by violating your dignity as a real human person—for example, by arbitrarily running you down with a truck just because he feels like crushing a solitary walker. This is a case that Thomson calls *Villainous Aggressor*.[284] It is crucial to note that even in such cases, it is permissible to kill only by using minimal lethal force, that is, the smallest amount of violence in a given context that is sufficient for being effective against an opponent, and that would also normally kill that opponent. If a villainous aggressor attacks you, and if as it so happens you also have an anti-tank gun with you, are skilled in using it, and are able to blow up the truck and stop him, and this is indeed the smallest amount of violence in this context that is sufficient for being effective in protecting yourself, and it would normally kill such an attacker, but the aggressor miraculously survives and is unconscious but seriously wounded, you are *not* then permitted, for example, to leave him lying in the street without calling for medical help, until he dies, to strangle him to death, or to torture him to death by waiting until he is conscious again and then using your knife to hack off many small parts of his body until he finally dies from this torture. This is because, as we saw in chapters 3 and 4 above, treating a real human person *as a mere thing* is absolutely always morally impermissible, and so is torturing.

This line of thinking also indirectly displays the moral rationale for permissibly killing villainous aggressors in self-defense. As we also saw in chapter 4, we are morally obligated, other things being equal, to prevent or reduce dignity-violating harms to real human persons. In this case, *I* am the real human person who is at risk of being mortally harmed by violating my dignity. So in this case, I am obviously *morally permitted* to try protect myself, by lethal force if necessary—although it must be minimal lethal force only—precisely because I am *morally obligated* to try to protect myself. Indeed, it is my personal deep moral responsibility to try to protect myself in such cases, as a matter of self-respect.

This indirectly raises two other points.

The first point concerns cases in which an agent is morally permitted to kill a villainous aggressor in self-defense, but simply is under-equipped, too vulnerable, or too weak, and thereby lacks the means, power, or wherewithal to fight off the villainous aggressor on her own. Now suppose that I am not the victim, but also that I do actually possess the lethal means, power, or wherewithal—may I permissibly kill the villainous aggressor on the victim's behalf? I will call this *killing in self-defense by proxy*. So the question is: Is killing a villainous aggressor in self-defense by proxy morally permissible or even morally

obligatory? Clearly and distinctly, *yes*, at the very least it is morally permissible, other things being equal. Moreover, in some particularly aggravated contexts—for example, when the villainous aggressor is the moral equivalent of a Nazi or an "ethnic cleanser"—it is even morally obligatory, on the assumption that the act is undertaken solely to prevent a dignity-violating harm to the villainous aggressor's victim, and also provided that I use only minimal lethal force.

The second point is something that I want to highlight for later discussion. By hypothesis in the self-defense cases we are considering, I am the *only* one who can save the relevant someone's life, namely, my own life. Obviously, this is not always true in cases of permissible life-saving. But in some cases, in part precisely and ineluctably contingently just because I am the *only* one who can save someone, it is my personal deep moral responsibility to do so, even if doing so will involve my sacrificing something of moral significance. It is easy enough to see how, in ordinary cases, I can be personally deeply morally responsible for saving my own life, even up to the point of killing a villainous aggressor in self-defense. But suppose that in fact I hate my own life and want to commit suicide? Should I then allow a villainous aggressor to kill me, without even lifting a finger?

No. Other things being equal, I should stop him if I can, even if I hate my own life, precisely because to do otherwise would be aiding and abetting a violation of my dignity as a real person. Correspondingly, other things being equal, I should not commit suicide except to end my own personhood-destroying suffering.[285] Hence merely hating my own life, or merely suffering intensely, is not sufficient to justify suicide. This is because at any time, no matter how awful and how miserable my life has been, as long as I am still a living, alert, and relatively sane higher-level or Kantian real human person, *I can always freely choose to change my life, and achieve principled authenticity, at least partially or to some degree*. As before, I am not saying either that this is in any way easy, or that I myself would ever actually be able to do it, but only that it is really possible, and that I ought to do it.

Back now to the original self-defense cases. We are assuming that, other things being equal, it is morally permissible to kill villainous aggressors in self-defense, provided that some other conditions are satisfied. But if the person who is a mortal threat to me is *not* villainous—that is, if the person who is a mortal threat to me is not also impermissibly threatening me—then is it still morally permissible for me to kill that person in self-defense, provided that some other conditions are satisfied? There are two different sorts of sub-case here.

The first sort of sub-case is what Thomson calls *Innocent Aggressor*,[286] which happens when some innocent bystander is forced or otherwise used by a villainous aggressor to do his life-threatening dirty work for him. For instance, a villainous aggressor who wants to kill me, and has the appropriate technology, or, for whatever reason, is in just the right causal circumstances, might use an innocent bystander as a human missile, aimed directly

at me—for example, he could push an innocent Fat Man off a high bridge down on top of me just as I walk on a narrow path beneath the bridge. Now let us suppose again, as per Thomson, that I just happen to have an anti-tank gun with me, am skilled in using it, and the only way I could save myself is by blowing up the Fat Man before he reaches me—would that be morally permissible?

Yes, it is morally permissible. Why? For the following reasons. If, contrary to hypothesis, the only way I could save myself were to duck or to jump out of the way, thereby letting the Fat Man die without trying to catch him or otherwise impede his fall, then that, surely, would be morally permissible. I am morally obligated, other things being equal, to prevent or reduce dignity-violating harms to real persons. In this case I am one of the real persons at risk of being mortally harmed by violation of his or her dignity (the unfortunate Fat Man, obviously, is the other). And moral obligation entails moral permissibility. So in this case, again, I am obviously morally permitted to protect myself, by lethal force if necessary—although it must be minimal lethal force only—precisely because I am morally obligated to protect myself. The further fact that in this case my minimal lethal force is turned upon an innocent bystander who unfortunately is being used by the villainous aggressor as a human missile aimed right at me, is morally irrelevant. By hypothesis, I cannot save the Fat Man. I can save only myself, and in this case I can save myself only by killing the Fat Man first. Other things being equal, I am obligated to protect myself from being harmed by being violated in my dignity as a real person. So, other things being equal, in the case in which the villainous aggressor uses an innocent bystander Fat Man as a human missile aimed right at me, I am morally obligated to kill the Fat Man in self-defense—provided, of course, that there is no other non-lethal way to save myself, and that I use only minimal lethal force. If other things are *not* equal, I might heroically and supererogatorily choose self-sacrifice. But in any case, obviously, I am morally *permitted* to kill the Fat Man in self-defense.

This brings us to the second and in fact crucial sub-case of killing innocent real persons in self-defense. It is "crucial" precisely because the fact that there is a mortally threatening innocent real human person in this case does not follow in any way from villainy, which, prima facie, might have seemed to be what was making the moral difference between morally permissible killing and morally impermissible killing. This is the case, following Thomson, I call *Innocent Attacker*.[287]

For instance, we can now imagine that there is no villainous aggressor in play and that the Fat Man has accidentally fallen off a bridge just as I am walking under it. So here he is now, in all his hugeness, accidentally hurtling down towards me, and the only way I can save myself is by blowing him up, provided, of course, that I happen to have brought my anti-tank gun with me again, am skilled in using it, and so-on. The Fat Man is completely innocent, and also in no way the puppet or tool of some villain. Would my killing him in self-defense still be morally permissible? And if so, then how can that be? This is what I call the *The Self-Defense Problem of Saving Lives*.

My answer to the first question, which I have already mentioned in section 5.2 above, is this. *Yes*, other things being equal, I am permitted to kill another real human person in self-defense, even if he is only an innocent attacker, provided that

> (i) killing is the *only* way I can protect myself from being mortally threatened in that context, and
> (ii) only minimal lethal force is used by me.

But what is my moral rationale for this claim? Here I can help myself to an argument I already formulated in section 3.4 above. This is what I wrote there:

> Consider a scenario in which I am a bicyclist and involved in a two-bicycle accident with another bicyclist, previously unknown to me (so: s/he is specifically *no*t a loved one, a close friend, or someone else I have explicitly or implicitly promised to aid or protect), that is no one's fault—for example, a sudden heavy gust of wind blows me and the other cyclist into one another. But unfortunately the accident happens on a busy street, and now the other cylist is lying unconscious on top of me, while suddenly a large Sport Utility Vehicle (SUV), being driven by a reckless college student, is barrelling directly towards both of us at high speed and is just a few yards away, unable to stop in time, or swerve so as to miss both of us. As it so happens, then, absolutely the *only* way I can save myself from being run over by the SUV is to push the unconscious other cyclist towards the speeding SUV, and roll sideways. The unconscious other cyclist is an innocent attacker in this case, and I hold that it would be morally permissible for me to kill him in the way I have described, other things being equal. The rationale is this. I am morally required, other things being equal, to provide benefits for real human persons, and also to prevent harm to them, *including* myself. Moreover, other things being equal, my untimely death is a bad and harmful thing for me. Also I am morally required, other things being equal, to pursue my own self-perfecting projects, which obviously will not be possible if I am dead. So self-defense is at the very least morally permissible, other things being equal, and is a first-order substantive ceteris paribus objective duty to myself. In this case, I am not treating the innocent attacker either as a mere means or as a mere thing, or with cruelty, and harming them by violating their dignity as a person—there is nothing "personal" in my pushing them off me in that way, thereby killing them. Indeed, if there were any other possible way I could push them off me, save myself, and *also* save their life, then I would do so. Nor am I being unkind specifically to *them*: I intend no cruelty whatsoever towards them. Moreover, I would also give my counterfactual rational consent to a scenario in *which I am killed in exactly the same way*, in a slightly different possible act-world in which our personal identities were switched, and unluckily *I was the unconscious cyclist*, and s/*he was the conscious cyclist accidentally pinned underneath me*. Therefore, in the actual world the unconscious cyclist's possible rational consent can be assumed, other things being equal, and I am also sufficiently treating them with respect and not violating their dignity, other things being equal—even though, obviously, I am seriously harming them by killing them.

There is no morally relevant difference between the case of the unconscious bicyclist who is an innocent mortal threat to me, and the case of the Fat Man who has villain-lessly and accidentally become a human missile aimed directly at me. They are both innocent attackers, and the only noticeable difference is that one is normal-sized and stationary on top of me, pinning me down, while the other is really huge and hurtling down towards me, and just about to crush me. But this noticeable phenomenological difference in proximity, distance, temporality, and causation is morally irrelevant in *this particular pair of cases*. Therefore, since pushing the unconscious cyclist towards the oncoming SUV is morally permissible, then blowing up the falling Fat Man must be morally permissible too.

It is quite instructive to note in passing, for later purposes of discussion in section 5.4, that although not every phenomenologically noticeable difference between cases is a morally relevant difference, nevertheless in some contexts/situations, various ineluctably contingent differences in proximity, distance, temporality, and causation, as well as egocentrically-centered emotional relations and social relations, are indeed morally relevant, and in fact partially determine the content of moral duties in those contexts. That is what The Morality De Re Principle says. But in *this particular pair of cases*, as it so happens, the phenomenologically noticeable differences in proximity, distance, temporality, and causation are every bit as morally irrelevant as the phenomenologically vivid fact that pushing an unconscious normal-sized cyclist towards a speeding SUV is not even close to being as spectacularly gory as blowing up a Fat Man with a handheld anti-tank gun.

Unfortunately, there is no simple algorithm for distinguishing between phenomenologically noticeable differences per se, and morally relevant differences. According to what I argued in chapter 2, the morally relevant differences are grounded on really existing moral principles in the Existential Kantian Ethics-based, No-Foolish-Consistency-approach-driven, nonideal Kantian hierarchical structuralist system of such principles, and cognitively accessible by means of rational intuition and/or careful reflection. Mere phenomenological noticeability, by contrast, does not ultimately hold up under these constraints. But these are only procedural guidelines, not a mechanical test.

There is one last variant on *Innocent Attacker* that I would like to consider. This is the case I will call *Well-Armed Innocent Attacker*.[288] Suppose that everything remains the same as in the human missile case of the Fat Man accidentally falling off a bridge onto me, and again I have brought my trusty anti-tank gun with me, am skilled in using, etc., *except that*

> (i) I am now a Fat Man too, and even huger than the original falling Fat Man, so that when he lands on me it will be, for him, just like landing on an air bag or in a foam pit and the original falling Fat Man's life will be saved although I will be killed, and
> (ii) the new falling Fat Man is armed with a gun, is skilled in using it, and can shoot me before I blow him up, thereby killing me but also saving his own life by landing on my even huger lifeless remains.

In other words, *Well-Armed Innocent Defensive Attacker* is a case in which the Fat Man and I are reciprocally self-defending innocent attackers. Would it be permissible for the armed falling Fat Man to shoot me before I blow him up, thereby saving his own life?[289] The answer seems to be clearly, *yes*, by another extension of the unconscious cyclist argument, provided that this is the only way the well-armed falling Fat Man can save himself, and also that he is using only the minimal lethal force needed. The further phenomenologically vivid fact that there are two possible self-defense scenarios built into the very same human missile scenario is, however, morally irrelevant.

I should add, before moving on, that all the semi-facetious talk of guns, anti-tank weapons, and so-on, in this section may be somewhat misleading, or even offensive, not only to others but also to *myself*, in other argument-contexts. That is because I also strongly hold that, other things being equal, the possession and use of guns and other similar weapons is rationally unjustified and immoral, and that, in turn, is because the primary function of guns and other similar weapons is coercion, and coercion is rationally unjustified and immoral.[290] So all that shooting and blowing-up in these examples is really only for expository convenience in the thought-experiments, and simply because Thomson's original examples employed them too. In fact, I despise guns and the Second Amendment to the US Constitution alike, with a moral and political passion.[291]

5.4 PONDS, VINTAGE SEDANS, AND ENVELOPES: THE FAMINE RELIEF PROBLEM

In studying The Trolley Problem of Saving Lives in section 5.2 above, we saw that on strictly non-consequentialist, and indeed Existential Kantian Ethics-based, grounds alone, other things being equal, sometimes it is morally permissible to kill a few innocent people in order to save more innocent people, although this is not always permissible. But in "Famine, Affluence, and Morality,"[292] Singer famously argues that it is morally obligatory for relatively well-off people like us *always to do whatever we can to save the lives of innocent people who are in great danger*, and in particular, that it is morally obligatory for relatively well-off people like us *always to do whatever we can to prevent the suffering and deaths of innocent people from famine anywhere in the world*. Singer's thesis of course expresses a robustly altruistic version of act utilitarianism. Here, in turn, is a rational reconstruction of Singer's famous argument.

Singer's Famine Relief Argument

(i) Suffering and death from famine are very bad.
(ii) Here is a candidate moral principle for adoption as a duty:

> *The Super-Strong Saving Others Principle*: If it is in our power to prevent something bad from happening to other people, without sacrificing anything of *comparable moral value*, then we ought, morally, always to do it.

(iii) Suppose that you do not agree with The Super-Strong Saving Others Principle. Then consider instead the following weaker moral principle:

> *The Strong Saving Others Principle*: If it is in our power to prevent something *very* bad from happening to other people, without thereby sacrificing anything *morally significant*, then we ought, morally, always to do it.

(iv) Both of the moral principles stated in (ii) and (iii) hold even if I am not the closest one to the endangered person, and even if I am not the only one who can save that person. In other words, neither the factors of *proximity* and *distance*, nor the factor of *uniquely effective aid*, has any moral relevance.

(v) Consider now the following example, *The Pond*:

> If I am walking past a shallow pond and see a child drowning in it, and as it happens I am the closest one to the child and also I am the only who can save the child, then I ought to wade in and pull the child out. This will mean getting my nice clothes muddy (and possibly ruining them), but this is morally insignificant, whereas by sharp contrast the death of the child would be a very bad thing, hence I am morally obligated to wade in and save the child.

(vi) Our strong commonsense moral intuitions about the shallow pond case confirm either The Super-Strong Saving Others Principle or The Strong Saving Others Principle.

(vii) The case of famine relief is precisely morally analogous to the shallow pond case in *Pond*, and further confirms either The Super-Strong Saving Others Principle or The Strong Saving Others Principle.

(viii) This entails that we accept the following moral principle as our duty:

> *The Strong Famine Relief Principle*: It is morally obligatory for relatively well-off people like us always to do whatever we can, short of sacrificing anything morally significant, to prevent the suffering and deaths of innocent people from famine anywhere in the world.

I think that everyone who carefully considers Singer's argument will agree that steps (i) and (v) are true. So that leaves steps (ii), (iii), (iv), (vi), (viii), and (viii) as possible targets for criticism.

Now consider the following pair of cases, due to Peter Unger, formulated in his words:[293]

> *The Vintage Sedan*. Not truly rich, your one luxury in life is a vintage Mercedes sedan that, with much time, attention, and money, you've restored to mint condition. In particular, you're pleased by the auto's fine leather seating. One day, you stop at the intersection of two small country roads, both lightly travelled. Hearing a voice screaming for help, you get out and see a man who's wounded and covered with a lot of his blood. Assuring yourself that his wound's confined to one of his legs, the man also informs you that he was a medical student for two full years. And despite his expulsion for cheating on his second year exams,

> which explains his indigent status since, he's knowledgeably tied his shirt near the wound so as to stop the flow. So, there's no urgent danger of losing his life, you're informed, but there's great danger of losing his limb. This can be prevented, however, if you drive him to a rural hospital fifty miles away. "How did the wound occur?" you ask. An avid bird watcher, he admits that he trespassed on a nearby field and, in carelessly leaving, cut himself on rusty barbed wire. Now, if you'd aid this trespasser, you must lay him across your fine back seat. But, then, your fine upholstery will be soaked through with blood, and restoring the car will cost over five thousand dollars. So, you drive away. Picked up the next day by another driver, he survives but loses the wounded leg.
>
> *The Envelope*. In your mailbox, there's something from (the US Committee for) UNICEF. After reading it through, you correctly believe that, unless you soon send in a check for $100, then, instead of living many more years, over thirty more children will die soon. But, you throw the material in your trash basket, including the convenient return envelope provided, you send nothing, and, instead of living many years, over thirty more children soon die than would have had you sent in the requested $100.[294]

Almost everyone who rationally considers these cases regards your action in Vintage Sedan as obviously morally impermissible and also your action in Envelope as obviously morally permissible. And this is even despite the highly disturbing and cognitively dissonant fact that not only is the loss of a leg in Vintage Sedan far less serious than the loss of thirty children's lives in Envelope, but also the cost of repairing your vintage sedan's upholstery ($5000) in Vintage Sedan is far greater than than the cost of donating to UNICEF ($100) in Envelope. What Unger argues is that our initial moral judgments in Vintage Sedan and Envelope are deeply erroneous and need to be revised.

This is what Unger calls a *Liberationist* solution to The Famine Relief Problem of Saving Lives, as opposed to a *Preservationist* solution that explains and justifies our initial moral judgments. In fact, he says, the injured leg emergency situation in Vintage Sedan is precisely morally analogous to the drowning emergency situation in Pond; and so too the famine relief emergency situation in Envelope is precisely morally analogous to the drowning emergency situation in Pond. Therefore, according to Unger, relatively well-off people like us ought always to give (for example) $100 to UNICEF (or OXFAM, CARE, etc.) whenever we can, and arguably we should also always be prepared to sacrifice a leg or to kill some other innocent person in order to save many faraway starving children whenever we can. And this of course is in perfect conformity with Singer's Famine Relief Argument.

I do completely agree with Unger and Singer that in *Pond* not only is The Strong Saving Others Principle validated, but also another principle I will call *The Surprisingly Strong Saving Others Principle:*

> In certain types of cases, if it is in our power to prevent something very bad from happening to other people, even when we will thereby have to sacrifice something morally significant,

> although without sacrificing anything of comparable moral value, then we ought, morally, always to do it.

What The Surprisingly Strong Saving Others Principle means in relation to *Pond* and all other *Pond*-type cases is this: Even if I were wearing, for example, a fabulously rare and valuable gold-plated Rolex watch, in which I had invested most of my life savings, but which would be completely ruined by my saving the drowning child, so that I thereby lost most of my life-savings, then I would still be morally obligated to save that child in that actual context. Tough luck for Bob.

But on the face of it, the Unger-Singer claim that we must validate The Surprisingly Strong Saving Others Principle in all *Envelope*-type cases, not only in all *Pond*-type cases, seems much too morally demanding in the sense that it mistakenly substitutes conduct that is in fact supererogatory, for conduct that is at most permissible and certainly not obligatory. Such superergatory conduct would include, for example, always giving $100 to UNICEF (or OXFAM, CARE, etc.) whenever we can, ruining the leather seats of my vintage sedan to the tune of $5000, destroying my beloved Bugatti to the tune of most of my life savings, or destroying my fabulously rare and valuable gold-plated Rolex watch to the same tune. Correspondingly, it clearly seems to be permissible for me, a relatively well-off individual, at least sometimes, to throw the envelope in the trash, even though I actually still have (for example) $100 that I could spend without sacrificing anything of moral significance, not to mention actually still having a Bugatti or gold-plated Rolex watch that I could afford to lose—although only just *barely* afford to lose.

As Kant notes, not only are we permitted to pursue our personal projects whenever this is consistent with the Categorical Imperative, but we also have the self-regarding duty, other things being equal, to develop our abilities and pursue a self-perfecting life-project:

> When it is said that it is in itself a duty for a man to make his end the perfection belonging to man as such (properly speaking, to humanity), this perfection must be put in what can result from his *deeds*, not in mere *gifts* for which he must be indebted to nature; for otherwise it would not be his duty. This duty can therefore consist in *cultivating* one's *capacities* (or natural predispositions), the highest of which is *understanding*, the capacity for concepts and so too for those concepts that have to do with duty. At the same time this duty includes the cultivation of one's *will* (moral cast of mind), so as to satisfy all the requirements of duty. (*MM* 6: 386-387, italics in the original)

> Adversity, pain, and want are great temptations to violate one's duty. It might therefore seem that prosperity, strength, health, and well-being in general, which check the influence of these, could also be considered ends that are duties, so that one has a duty to promote *one's own* happiness and not just the happiness of others. But then the end is not [merely] the subject's happiness but his morality, and happiness is merely a means for removing obstacles to his morality—a permitted means, since no one else has a right to require of me that I sacrifice my ends if these are not immoral. To seek prosperity for its own sake is not

> directly a duty, but indirectly it can well be a duty, that of warding off poverty insofar as this is a great temptation to vice. But then it is not [merely] my happiness but the preservation of my moral integrity that is my end and also my duty. (*MM* 6: 388, italics in the original)

Obeying the self-regarding duty, other things being equal, to develop ourselves and pursue a self-perfecting life-project of course at least sometimes requires spending money I could give to famine relief instead—for example, buying myself some philosophy books and/or some DVDs or a subscription to *Filmstruck*, or taking the occasional, short, modestly-expensive vacation from non-stop work on THE RATIONAL HUMAN CONDITION and other philosophical projects, not to mention sometimes buying some food and drink that is slightly more expensive than other merely adequately nutritious fare—without sacrificing anything of comparable moral value to me or indeed of any moral significance to me.

So what I will call *The Famine Relief Problem of Saving Lives* is this: How can we *accept* steps (i) and (v) of Singer's Famine Relief Argument, while also *rejecting* steps (ii), (iii), (iv), (vi), (vii), and (viii) of it, and *also reject* Unger's conclusion, which entails validating The Surprisingly Strong Saving Others Principle in all *Envelope*-type cases? Or otherwise put, what is the morally relevant difference between *Pond* and *Envelope*?

As against both Singer and Unger alike, my view, The Morality De Re Solution to The Famine Relief Problem of Saving Lives, is that the ineluctably contingent factors of proximity, distance, temporality, and causation, as well as egocentrically-centered emotional relations and social relations, and *uniquely effective aid* are morally relevant in these cases, insofar as the following first-order substantive ceteris paribus objective moral principle is both true and also applies directly to *Pond* and all other *Pond*-type cases—

> The First Morality De Re Saving Others Principle:
>
> If it is in my power to prevent something very bad from happening to another real human person, whenever I am the closest one, among all the close ones,[295] to the endangered person, and also I am the only one who can save that person, without sacrificing anything of moral significance to me, then I ought, morally, always to do it, other things being equal. And I should also always do it even if it means sacrificing something of moral significance to me. The ineluctably contingent factors of distance, proximity, temporality, and causation, as well as egocentrically-centered emotional relations and social relations, combine to make saving that real person my personal deep moral responsibility in that context. Even so, this principle is my duty only on condition that I am not morally required to repeat this act of sacrifice to the point at which it undermines my obligatory life-project, other things being equal, of developing my abilities and perfecting myself.

And here is the rationale for that principle. If I let the drowning child die in *Pond* or if I let the injured bird watcher lose his leg in *Vintage Sedan*, then *not only* would I be regarding

and treating the child and bird-watcher as mere things, as worth less than my clothing and my leather upholstery, which is a direct violation of The Formula of Humanity as an End-in-Itself, *but also* I would be treating someone without their actual or possible rational consent. In these two contexts, it is manifestly "up to me" to prevent harm to the child and bird-watcher. Various ineluctably contingent features of the context have conspired to make it my personal deep moral responsibility to help them, even if the cost to me of preventing harm to them is suprisingly high.

This rationale also directly supports Singer's basic intuition about the validation of The Surprisingly Strong Saving Others Principle in *Pond*-type cases, which he pumps in another paper by means of the story of Bob, the foreign sportscar enthusiast—already recounted in the second epigraph for this chapter—who must choose between, on the one hand, saving an innocent child from being run over by a train, and on the other, saving his beloved Bugatti in which he has invested most of his life savings:

> When Bob first grasped the dilemma that faced him as he stood by that railway switch, he must have thought how extraordinarily unlucky he was to be placed in a situation in which he must choose between the life of an innocent child and the sacrifice of most of his savings. But he was not unlucky at all. We are all in that situation.[296]

Singer is right that Bob is not "unlucky" at all—in the normal sense of that word. But at the same time, Bob's situation expresses *the ineluctable contingency of moral essential indexicality*, which, as we saw above, is a species of moral luck.[297] So Singer is bang-on correct about Bob. But at the same time Singer is off-target and mistaken in claiming *that we are all in that Bob-like situation*, in all cases in which we can prevent mortal harm to others. It is *Bob's* personal deep moral responsibility to save the child even at the cost of his beloved Bugatti, but not *our* personal deep moral responsibility—at least not in *that* situation, although it may well be our personal responsiblity in other situations of that type, depending on our moral luck.

Correspondingly, The First Morality De Re Saving Others Principle tells us that it is not *my* personal deep moral responsibility to give money in *Envelope*—even despite my also being named "Bob," aka Bugatti-less Bob—because I am neither the closest one, among all the close ones, to the endangered children, nor am I the only one who can save them. So that would smoothly explain the moral difference between *Pond, Vintage Sedan*, and *Bob's Bugatti* on the one hand, and *Envelope* on the other. And this in turn provides a Preservationist solution to The Famine Relief Problem of Saving Persons.

It should be particularly noticed, moreover, that The First Morality De Re Saving Others Principle is also strong enough *to override a person's right to control his or her own body*. If you were, for example, to modify slightly a case described by Thomson in "A Defense of Abortion," *sick unto death*, and needed only Jane Fonda's cool touch on your fevered brow in order to save your life,[298] and by pure chance Ms. Fonda was right there

in the same room with you, then it would be morally impermissible for her to refuse the laying-on of hands, even though she had never either explicitly or implicitly promised or agreed to provide you with what you needed in order to survive. Again, according to The First Morality De Re Saving Others Principle, it would simply be Ms. Fonda's personal deep moral responsibility to touch you in that context.

But on the other hand, if Ms. Fonda were in Los Angeles and you were in Winnipeg, Manitoba, Canada, and if Your People were in touch with Her People, then she would not be morally required to fly out and provide that aid and succor—as morally heroic and sublime, of course, as it would be for her to do so. Similarly but conversely, if you are in Winnipeg, then you are not morally required to provide a painless pint of blood for an innocent stranger living in Los Angeles who will die if you do not provide that pint, even if you are in fact the only person who has the special type of blood the stranger needs in order to survive. In the same situation, a great many people no doubt *actually* would give their blood to the endangered stranger in L. A., which would truly be morally heroic and sublime of them—in the sense of fully keeping rational faith with the Categorical Imperative—although also supererogatory. So you do not morally have to do as they do. We are not generally morally required, other things being equal, to provide the means of survival for other real persons when we have not promised to do this or otherwise committed ourselves to protecting them from mortal harm. We are only generally morally required, other things being equal, to prevent or reduce harms that violate the dignity of real persons.

Therefore, the fact that I am the only one who can save some other mortally threatened innocent real human person's life is not alone sufficient to entail an obligation for me to to save that stranger's life. It is only if I am also the closest one, among all the close ones, to that mortally threatened innocent stranger, and only if I am also the only one who can save the innocent stranger, and only if I am also not also required to repeat this act of sacrifice to an unreasonable life-changing extent—even though it might require that I sacrifice something of moral significance to me, not to mention my sacrificing something that is of moral value to me but not of any moral significance to me, depending on my moral luck—that it will be my personal deep moral responsibility to save the innocent stranger. To require more than this of me, however, would be too morally demanding.

The explicit mention of strangers in the last example raises an extremely important issue that has been lurking in the wings of the entire discussion up to this point, namely *The Problem of (Im)Partiality.* I have been assuming in all the saving-lives cases under consideration that none of the people whose lives are to be saved are either loved ones, friends, or family. But what if, holding everything else fixed, there are *two* mortally threatened innocent people, and one of them is a loved one, friend, or family member? Am I morally justified in choosing to save that beloved, or in any case special, person over saving the stranger, and if so, how can I do so without violating the moral demand for impartiality?

The answer is that emotional relations and social relations are among the indexical, egocentrically-centered factors to which The Morality De Re Thesis is directly sensitive. Hence, in context, it becomes my personal deep moral responsibility to save my loved one, friend, or family member ahead of my saving any stranger. For example—and being fully mindful here of avoiding an *ad hominem* ("attack the person instead of his argument") or *tu quoque* ("look who's talking") fallacy—as a matter of actual fact, Peter Singer himself spent a substantial amount of money on special nurses for his mother, who was suffering from advanced Alzheimer's Disease, that he could have used to save many other innocent mortally threatened strangers from dying.[299] And clearly *he was right to do so,* even if it was inconsistent with his own moral principles. Moreover, the manifest fact that the contextual factors of special emotional relations and social relations between me, or Singer, and one of the mortally threatened people, partially determines this result, is *not in and of itself* a matter of egoism or self-interest, even if it happens to *coincide with* self-interest. Or otherwise put: *egocentric centering* is *not* the same as psychological or ethical *egoism*.

There may seem to be some critical wiggle-room for the Singer-style act-utilitarian here, who can hold that if, in a certain context, the psychological costs of not acting on the basis of the bonds of love, friendship, or family are such that it would substantially affect the ability of the agent to act altruistically in the future, then s/he should act on the basis of these bonds in that context. But that is mistaken. Even if Singer had been angry at his mother, or callously indifferent to her suffering, or for any other reason emotionally distanced from her, *he still ought to have done what he actually did*. This, in turn, can be clearly seen in the manifest fact that even if I happen to be angry at, or callously indifferent to, or for any other reason emotionally distanced from, my loved one, friend, or family member, in some context, *nevertheless I am still obligated in that context to save them ahead of the stranger*. The impartiality of morality thus necessarily *includes* The Morality De Re Thesis.

Now suppose that, in any given context, The First Morality De Re Saving Others Principle has made it non-obligatory for me to prevent the suffering and death of some mortally threatened far away innocent strangers. Nevertheless, it would be morally wrong for me, a relatively well-off individual, *never to try to do anything whatsoever* to prevent the suffering and death of innocent people even when they are far away and I am not the only one who can help them, and it will not involve my sacrificing anything morally significant, and I am not being morally required to iterate my act of sacrifice to an unreasonable life-changing extent. This point can be made in two steps.

First, let us suppose that it is true that in cases in which it really is my personal deep moral responsibility to save someone, then I am also not required to iterate this act of sacrifice to an unreasonable life-changing extent. Nevertheless, it would still be entirely morally permissible, and indeed morally heroic and sublime, for me to undertake repeated acts of sacrifice as proper parts of my obligatory life-project, other things being equal, of developing my abilities and perfecting myself. In other words, it would be entirely morally

permissible and indeed morally heroic and sublime for me to change my life and become a serious moral activist and life-saver, as a proper part of my unique life-project in pursuit of the realization of principled authenticity, at least partially or to some extent. But it is also entirely morally permissible for me to do *something else* with my life instead, that is equally in pursuit of of the realization of principled authenticity, at least partially or to some extent.[300]

Second, it is clearly morally wrong for me not always to inform the proper authorities of great danger to others when I have good reason to believe that the proper authorities do not know about the great danger and it requires only a minimal effort to inform them—for example, by calling them on my smart phone—and I am also not thereby required to iterate this sort of informing act to an unreasonable life-changing extent. Therefore, even when I am not the closest one, among all the close ones, to some mortally endangered innocent stranger, and even when I am not the only one who can save that stranger, nevertheless I can still have some life-saving-relevant duties towards that stranger.

Now if, first, it is the case that it is fully permissible and also morally great and heroic for me to choose to become a serious moral activist and life-saver, and if also, second, I am morally obligated always to inform the authorities of great danger to others even when I am not the closest one, among all the close ones, or the only one who can save them, and it requires nothing of moral significance for me to do this, and I am not required to repeat this to an unreasonable extent, then the following further moral principle is also clearly true—

> The Second Morality De Re Saving Others Principle:
>
> If it is in my power to prevent something very bad from happening to other people, living anywhere in the world, without thereby sacrificing anything morally significant, then I ought, morally, sometimes to prevent the very bad thing from happening, even if I am neither the closest one, among all the close ones, to the endangered people, nor the only one who can save them, as long as I am not required to repeat this act of sacrifice to an unreasonable life-changing extent.

And here is the rationale for this principle. Suppose that all relatively well-off people like us never do anything whatsoever to prevent very bad things from happening to other people whenever we are neither the closest ones, amongst all the close ones, to the endangered people nor the only ones who could save them, and it does not involve our sacrificing anything morally significant, and we are not required to iterate this act of sacrifice to an unreasonable life-changing extent. In short, suppose that I never take any personal deep moral responsibility whatsoever to prevent harm to anyone, whenever this is not strictly morally imposed on me. Then what?

Well, *here's what*, in three steps.

First, this refusal to take personal deep moral responsibility entails that the world would be a morally much worse place than it would be if we at least sometimes saved others who are in great danger even when it is not morally obligatory to do so. And of course, according to Existential Kantian Ethics, we do indeed have a duty, other things being equal, to promote the happiness of other real persons either by preventing or reducing dignity-violating harm to them or by producing positive benefits for them.

Second, if I never take any personal deep moral responsibility whatsoever for anyone else's well being whenever it is not strictly morally forced on me by the context or by my antecedent promises, then surely that is *morally awful* of me, in the sense of being significantly out of spirit with the Categorical Imperative, although not strictly morally impermissible.

Third, and perhaps most importantly, my failure ever to take personal deep moral responsibility for saving others whenever this is not strictly morally imposed on me, would surely entail my regarding all those greatly endangered people as mere things, worth less than even the most trivial amount of hassle to myself, even if I am not thereby strictly speaking treating them as mere things. Now not only is it morally impermissible to *treat* people as mere things, it is also morally awful of me to *regard* them as mere things. This is because regarding people as mere things is a constitutively necessary condition of treating people as mere things, from which anyone's wicked intention to treat people as mere things would naturally flow, even if it is neither a strictly or logically necessary condition nor a sufficient condition.

This last point also needs a few follow-up points, in order to avoid possible misunderstandings. It is fully conceivable and therefore really possible that someone could treat other people as mere things without *self-consciously* regarding them as mere things, perhaps through some sort of process of serious self-deception—like the fanatical utilitarian do-gooder, Mr Gradgrind, in Dicken's *Hard Times*, who is always in fact hurting other people for the sake of the "general good." And it is also fully conceivable and really possible to regard other real persons as mere things while only *rarel*y treating them as mere things, if the opportunity rarely arises, or the fear of being caught or reprisal somehow usually suppresses the desire to degrade others. But in this thoroughly nonideal natural and social world, direct engagement with other real human persons is not merely really possible, but instead actual and ubiquitous. Therefore, it is not merely a logical possibility or an option, but instead just an ineluctably contingent fact about our real personal lives that we are always rubbing elbows with other people and always bumping up against them, like riders in a crowded subway. Hence we actually always *do* treat other people in some way or another. Then in this thoroughly nonideal world, actually regarding other people as mere things is *naturally poised for treating them as mere things*. It is psychologically unrealistic in the extreme to think that someone who is so hideously callous, or so titanically selfish, that he really and truly regards everyone else as a mere thing, would *never* act on the basis of this near-satanic way of regarding other people.

Therefore, for all those reasons, The Second Morality De Re Saving Others Principle is true.

5.5 CONCLUSION

Other things being equal, it is sometimes morally permissible for us to kill a few innocent people in order to save significantly more innocent people from dying. But, other things being equal, we are not morally permitted to force innocent bystanders into mortally threatening situations and kill them in order to save significantly more innocent people from dying. It is also sometimes morally permissible, other things being equal, to kill completely innocent attackers in self-defense. Still, other things being equal, it is obligatory to do whatever we can, even though we must sacrifice something of moral significance, not to mention sacrificing things that are of moral value but not of any moral significance, in order to save mortally endangered innocent strangers whenever we are the closest ones, among all the close ones, and the only ones who can save them, and it is not required of us that we repeat this act of self-sacrifice to an unreasonable life-changing extent. It is in fact our personal deep moral responsibility to prevent harm to strangers in those contexts. But if we are either not the closest ones, among all the close ones, or not the only ones who can save those mortally endangered strangers, then, other things being equal, while it remains fully morally permissible for us always to do whatever we can to save them—even while giving up things of some moral significance or comparable moral significance, and indeed even while it is morally great and heroic to become a serious moral activist and life-saver—nevertheless it is not morally obligatory for us always to save them. Yet we are morally obligated sometimes to save them, provided that we are not required to give up anything of moral significance in so doing, and that we are not required to repeat this act of sacrifice to any unreasonable life-changing extent, even if we are not the closest ones, among all the close ones, and even if we are not the only ones who can save them, other things being equal. Otherwise, we would be regarding other real persons as mere things, which, in this thoroughly nonideal natural and social world, is naturally poised for treating them as mere things, and therefore it is morally impermissible.

Commonsense moral intuition generally supports these principles. But that is not the reason, or at least not the *fundamental* reason, why these principles are correct. These principles are correct, fundamentally because they flow from the No-Foolish-Consistency-approach-driven, nonideal Kantian hierarchical structuralist theory of moral principles. Indeed, all of these commonsensically intuitive moral principles are ultimately explained and justified by The Formula of Humanity as an End-in-Itself, together with other first-order substantive ceteris paribus other-regarding and self-regarding objective moral principles deriving from LEVEL 2 of the hierarchy of principles, together with the three basic side-constraints or side-principles on the hierarchy (namely, No-Global-Violation,

Excluded Middle, and Lesser Evil)—see chapter 2 above—together with morally essentially indexical factors, including those that collaterally determine an effective solution to The (Im)Partiality Problem.

I conclude that The Morality De Re solutions to The Trolley Problem of Saving Lives, The Self-Defense Problem of Saving Lives, and The Famine Relief Problem of Saving Lives are each intelligible and defensible, and also that when they are conjoined, they provide an adequate unified solution to the general moral Problem of Saving Lives and The (Im)Partiality Problem alike. So the moral Preservationism that flows from Existential Kantian Ethics is correct, and Unger's and Singer's moral Liberationism, which flows from act consequentialism, is mistaken.

And look: There are those all-too-familiar envelopes from CARE, OXFAM, UNICEF, *Doctors Without Borders*, etc., etc., sitting on my kitchen counter, or on your kitchen counter. *If not now, then when?*

Chapter 6

RAGE AGAINST THE DYING OF THE LIGHT: THE MORALITY OF ONE'S OWN DEATH

Accustom yourself to believing that death is nothing to us, for good and evil imply the capacity for sensation, and death is the privation of all sentience; therefore a correct understanding that death is nothing to us makes the mortality of life enjoyable, not by adding to life a limitless time, but by taking away the yearning after immortality. For life has no terrors for him who has thoroughly understood that there are no terrors for him in ceasing to live. Foolish, therefore, is the man who says that he fears death, not because it will pain when it comes, but because it pains in the prospect. Whatever causes no annoyance when it is present, causes only a groundless pain in the expectation. Death, therefore, the most awful of evils, is nothing to us, seeing that, when we are, death is not come, and, when death is come, we are not. It is nothing, then, either to the living or to the dead, for with the living it is not and the dead exist no longer.[301]

Look back at time … before our birth. In this way Nature holds before our eyes the mirror of our future after death. Is this so grim, so gloomy?[302]

The greatest stress. How, if some day or night a demon were to sneak into your loneliest loneliness and say to you: "This life as you now live it and have lived it, you will have to live once more and innumerable times more; and there will be nothing new in it, but every pain and every joy and every thought and sigh and everything immeasurably small or great in your life must return to you—all in the same succession and sequence…" Would you throw yourself down and gnash your teeth and curse the demon who spoke thus? Or did you once experience a tremendous moment when you would have answered him: "You are a god, and never have I heard anything more godly." If this thought were to gain possession of you, it would change you, as you are, or perhaps crush you. The question in each and every thing, "Do you want this once more and innumerable times more?" would weigh upon your actions as the greatest stress. Or how well disposed would you have to become to yourself and to life to *crave nothing more fervently* than this ultimate eternal confirmation and seal?[303]

6.431 [I]n death … the world does not change, but ceases.

> 6.4311 Death is not an event of life. Death is not lived through. If by eternity is understood not endless temporal duration but timelessness, then he lives eternally who lives in the present. Our life is endless in the way that the visual field is without limit.[304]

> Grave men, near death, who see with blinding sight
> Blind eyes could blaze like meteors and be gay,
> Rage, rage against the dying of the light.[305]

6.1 INTRODUCTION

The image featured at the front of this book, and also at the fronts of volumes 1, 3, and 4 of THE RATIONAL HUMAN CONDITION, is of a painting by Thomas Whitaker, "The Human Condition." Whitaker was on death row for 11 years; and his death sentence was commuted to life imprisonment less than an hour before his scheduled execution on 22 February 2018. Leaving aside, for the time being, the morality of social-political questions about crime-&-punishment,[306] Whitaker's moral-existential predicament poignantly raises the following question:

> "Is a rational human life worth living?"

My first, short answer is: *Yes, even despite all the natural and moral evil, pain, and suffering that tend to support a contrary answer*.

> "Then *why* is a rational human life worth living?"

My second, longer answer is that a rational human life is worth living, even despite all the natural and moral evil, pain, and suffering that tend to support the contrary, precisely because of the opportunities it provides us for the achievement or realization of *principled authenticity*, at least partially or to some degree. In a nutshell, principled authenticity is an existentially extended and reformulated version of what Kant calls a *good will*, in that it is a coherent fusion of what he calls *autonomy* together with what Kierkegaard calls *purity of heart*. As I am construing these, autonomy is a rational minded animal's capacity for deep freedom, or up-to-me-ness, and moral self-legislation; and purity of heart is psychological coherence, single-mindedness, and wholeheartedness. And it is an essential feature of the rational human condition that we do not live ideally, in a void, or alone. The world we live in, is essentially *a thoroughly nonideal natural and social world*. So I am saying that a rational human life is worth living precisely because of the opportunities it provides us for incarnating autonomy and purity of heart, in solidarity with all other real human persons and alongside all other minded animals, everwhere, in this thoroughly nonideal natural and social world.

Now it is a brute fact of human life that we are always getting closer to what Kant aptly called "the end of all things" (*EAT* 8: 327-339), whether this will be a purely natural ending to everything human, or a man-made Apocalypse, like something out of Neville Shute's grim 1957 novel, *On the Beach*. But at a first-person level, it is also a brute fact that from the very moment I begin to live as the conscious subject of my own real personal life, I am always getting closer to the cessation or end of that life. Therefore, I am always getting closer to my own death. In that sense, my life-process is identically the same as the process of my dying. My own life is also my own death. This recognition, as they say, concentrates the attention.

What does one's own death mean, both in itself and also morally speaking? In this chapter, I will argue for a doctrine I call *The Rage-Against-the-Dying-of-the-Light Theory*, first, by working out an account of the nature and moral value of one's own death, and then, second, by considering the various first-order substantive ceteris paribus objective moral principles that govern five basic ways in which the process of my dying can happen:

(i) euthanasia,
(ii) self-sacrifice,
(iii) suicide,
(iv) accidental death, and
(v) natural death.

6.2 THE AMBIGUITY OF "DEATH"

What is death? Minimally, the English natural-language term "death," and correspondingly the concept of death, mean "the cessation or end of life." But unfortunately for those of us who live and die, and are also conscious and self-conscious, therefore able to think about our own lives and deaths—that is, all rational minded human animals, especially including all higher-level or Kantian real human persons—the concept of death is crucially ambiguous, in at least five different ways.

The first crucial ambiguity about the concept of death concerns *the type of life* we are talking about when we say that life ceases or ends:

(i) inorganic life,
(ii) organic life,
(iii) minded animal life, and
(iv) real personal life.

Correspondingly, there are four different sub-types of death:

(i*) inorganic death,

(ii*) organic death,
(iii*) minded animal death, and
(iv*) real personal death.

Many things have inorganic lives. This includes artificial or humanly-fabricated machines like automobiles, dishwashers, and refrigerators—indeed, these are often sold along with a legally binding "lifetime warranty"—but also more or less large scale non-artificial natural mechanisms like weather systems, tropical storms, mountains, mountain ranges, planets, stars, and galaxies. Principles of complex systems dynamics and evolutionary theory apply to their cosmic emergence, development, and eventual destruction. Such things therefore all encounter inorganic deaths at the end or cessation of their inorganic lives. Indeed, even the universe as a whole can, at least in principle, have an ultimate inorganic "heat-death," via entropy. In this sense, death is the cessation or end of something's characteristic mechanical operations or, more generally, the cessation or end of its inorganic complex thermodynamics.

Let us suppose that organic activity, as not only complex thermodynamic, but also self-organizing, hence purposive and self-guiding, and minimally spontaneous, hence underdetermined by what has preceded it and creative or productive, is not only epistemically or conceptually distinct, but also metaphysically distinct, from the activity of natural mechanisms.[307] As Kant compactly puts it, "a mere machine … has only a motive power, while the organized being possesses in itself a formative power" (*CPJ* 5: 374). Then all organisms have categorically and specifically *organic lives*, including micro-organisms, plants, and animals. It is not inconceivable that there could even be entire planets possessing organic lives, like the one imagined in Stanislaw Lem's brilliant science fiction novel *Solaris*, and represented visually in Andrei Tarkovsky's equally brilliant science fiction film *Solaris*, the eponymous Solaris. In any case, all *minded animals*, as living organisms, have categorically and specifically organic lives. And since all real human persons are also minded animals, so too do all real human persons have such organic lives. But, obviously, not everything that has an organic life—say, a unicellular micro-organism, or a plant—has either a minded animal life or a real human personal life.

So there is an important difference between, on the one hand, the cessation or end of an organic life, per se, and on the other hand, the cessation or end of either a minded animal life or a real human personal life. In particular, the real human personal life of a creature can temporarily or permanently cease or end, while its organic life or minded animal life continues:

(i) temporarily, for example, in cases of fainting, unconsciousness, or a coma;
(ii) permanently in one sense, while organic life but not minded animal life continues, for example, in cases of persistent vegetative states produced by an artificially-induced or disease-based brain-trauma, as in the famous Karen Ann Quinlan, Nancy Cruzan, and Terry Schiavo cases; and

(iii) permanently in another sense, while organic life and minded animal life both continue, for example, in cases of degenerative diseases like Alzheimer's, as in the famous case of the philosopher Iris Murdoch.

And on the other hand, at least in principle, real human personal life can continue across even very long temporary gaps in organic life and minded animal life, for example, in cryogenic re-animation.

Inorganic death, organic death, and the death of the minded animal, while philosophically important for various reasons, and by no means irrelevant to morality, nevertheless are not of *primary* moral importance. Only the deaths of *real persons* are of primary moral importance. Moreover, at the very center of The Web of Mortality are the lives and deaths of rational *human* minded animals and real *human* persons. Furthermore, and even more radically, as I mentioned in chapter 1, I hold that all meanings, truths, reasons, principles, and values of any kind, whether merely relative values or absolute values, are in the world *just because* rational human minded animals or real human persons *are* in the real world—or at least *just because* rational minded animals or real persons *really can be* in the world. Let us call this *the real-human-person-centered metaphysics of moral value*.

Now of all the minded animals and real persons we know, we ourselves are the only ones we have encountered, so far, that are also higher-level or *Kantian* human minded animals. I am talking about precisely the sort of human minded animals that are capable of actively reading and understanding these sentences, that are self-conscious, and who in turn are precisely those human minded animals that are also capable of reflecting on the meaning of their own lives. Insofar as the real-human-person-centered metaphysics of value is true, then since higher-level or Kantian human minded animals are at the very center of the class of real human persons, it follows that *real human persons like us* are at the very center of everything that really and truly matters: in that special metaphysical sense, *the Universe revolves around us*. By sharp contrast, as regards caring, value, and what really and truly matters, the Universe has *no* point of view.[308]

So for the specific purposes of Existential Kantian Ethics, in this chapter I will focus exclusively on the deaths of higher-level or Kantian human minded animals—*real human persons like us*. The deaths of such real human persons are categorically distinct from organic deaths per se and also from the deaths of minded animals per se—in the dual sense

(i) that both organic life and minded animal life can continue even though the life of the real human person like us has permanently ceased or ended, and
(ii) that both organic life and minded animal life can temporarily end or cease without the death of the real human person like us—even though, of course, every real human person is necessarily also a living organism and a minded animal.

In other words, in this chapter I am going to work out *the "us"-centered metaphysics of the moral value of death*.

The second crucial ambiguity about the concept of death concerns *the temporal duration* of the cessation or end of life, and in particular, whether it is

(i) temporary, or
(ii) permanent.

Now it is obvious that there can be temporary cessations or endings of rational consciousness—for example, fainting, unconsciousness, or a coma—that are not also permanent. Correspondingly if, as seems easily conceivable, were the technology and science of *cryogenics* to be developed somewhat further, then there could be even very long temporary cessations or ends of the organic lives of real human persons—the temporary deaths of their living bodies—that are neither the permanent deaths of their minded animals nor the permanent deaths of the higher-level or Kantian real human persons they are. For in these easily conceivable scenarios, when the body of the dead real human person is reanimated, then the real human person's life is also resumed, just as it would be after a fainting fit, unconsciousness, or coma. What seems far less easily conceivable is the supposed possibility of *reincarnation*, that is, the possibility of a higher-level or Kantian real human person's body's suffering a permanent organic death, therefore also being temporarily dead as a real human person, but then resuming their real human personal life in a new body. According to the Minded Animalist theory of the nature of personhood and personal identity that I work out and defend in *Deep Freedom and Real Persons*, chapters 6-7, and briefly sketched again in chapter 3 above, reincarnation is strongly metaphysically (and more precisely, synthetic a priori) impossible. This is because preserving the diachronic identity of a minded human animal's living body is a constitutively necessary condition of real human personal identity.[309] But in order to keep things relatively simple, I do not want to re-argue these somewhat controversal claims now; so for the purposes of this chapter, I will simply bracket any further discussion of reincarnation. In any case, the basic point I am making here is secured by the real possibility of *reanimation*.

Again for the purposes of this chapter, I am going to concentrate almost exclusively on the *permanent* deaths of real human persons like us; that is, I am going to concentrate almost exclusively on the annihilation or extinction of any such person as a rational, conscious, and self-conscious subject, forever. I say "almost exclusively," because later in this chapter I will critically consider the concept of immortality, or more precisely, the concept of an sempiternally endless or infinite higher-level or Kantian real human personal life. But aside from that discussion, and unless otherwise specified, I will be talking only about the permanent deaths of real human persons like us.

The third crucial ambiguity about the concept of death is in many ways the most important one. This concerns the moral-metaphysical distinction between

> (i) the *state* of my actually being dead (which I will call "death$_s$"), and
> (ii) the *process* of my dying (which I will call "death$_p$").

The state of my actually being dead, my death$_s$, necessarily occurs immediately *after* the process of my dying, my death$_p$. Now since I am concentrating almost exclusively on the permanent deaths of higher-level or Kantian real human persons, then my kind of death$_s$, once it has occurred, lasts forever. The process of my dying, my death$_p$, by sharp contrast, necessarily occurs *during* my life as a real human person. Otherwise put, death$_p$ is necessarily *infra*-life, whereas death$_s$ is necessarily *post*-life.

Many serious philosophical, existential, and moral confusions have been created by failing to distinguish between death$_s$ and death$_p$. For example, Lucretius asserted that

> since (i) the time prior to the beginning of my life and the time after the permanent cessation or end of my life are perfectly symmetrical and in effect metaphysical mirrors of one another, and
> since (ii) we are never (or at least almost never) concerned about the fact that we did not exist before we were born,
> then (iii) we should not be concerned about the time after we die, that is, we have no good reason to fear our own deaths.

But Lucretius was simply *wrong* about the symmetry or mirroring thesis, so his argument is unsound. The pre-natal non-existence of a higher-level or Kantian real human person is *essentially* different from her death$_s$, precisely because her death$_s$ is *necessarily post-life*, and therefore it *inherently presupposes her actual death$_p$*, whereas her pre-natal non-existence is necessarily *not* post-life, and therefore it does *not* moral-metaphysically include her actual death$_p$.

But that is by no means the worst of the confusions that have been created by failing to distinguish between death$_s$ and death$_p$. As we shall see later in the chapter, the participants in some of the leading recent and contemporary philosophical discussions of the nature of death have consistently failed to draw the distinction between the state of actually being dead and the process of dying, and have therefore fallen into serious confusions about whether death is a always a bad thing for the one who died, or not. Sometimes they are talking about death$_s$; sometimes they are talking about death$_p$; and sometimes it is crucially unclear precisely *which* kind of death they are talking about. In any case, as we will also see, it is entirely possible and perfectly coherent to hold

> (i) that a higher-level or Kantian real human person's $death_s$, by its very nature, is necessarily *neither* a good thing *nor* a bad thing for the one who dies (hence never a good thing and never a bad thing for the one who dies),

while at the same time also holding

> (ii) that a higher-level or Kantian real human person's $death_p$, by its very nature, is *sometimes* a good thing for the one who dies and also *sometimes* a bad thing for the one who dies.

These points lead on naturally to the fourth crucial ambiguity about the concept of death. This concerns the fact that a higher-level or Kantian real person's permanent death, whether this is her $death_s$ or her $death_p$, can be considered and/or evaluated

> either (i) from *the inside*, that is, from the first-person point of view,
> or (ii) from *the outside*, that is, from the third-person point of view.

Following David Suits, who originally discovered this deeply important distinction—or in any case, who first formulated it clearly[310]—I will say that whenever a higher-level or Kantian real human person's $death_s$ or her $death_p$ is considered and/or evaluated from the first-person point of view, then this is considering or evaluating some fact that is *for* the one who died, and therefore an *intrinsic* or internal fact with respect to that higher-level or Kantian real human person. But by sharp contrast, whenever a higher-level or Kantian real human person's $death_s$ or her $death_p$ is considered or evaluated from the third-person point of view, then this is considering or evaluating some fact that is only *about* the one who died, and therefore at best an *extrinsic* or external fact with respect to that higher-level or Kantian real human person.

The main reason this distinction is so important, as we will see, is that although a higher-level or Kantian real person's $death_s$ or $death_p$ can involve various good or bad facts *about* her, from the third-person point of view, it does *not* follow that any of these facts is a good or bad fact *for* her. So apart from Suits, few philosophers who have discussed the nature of death have been able to recognize that although there may be good arguments showing that the permanent $death_s$ of a higher-level or Kantian real human person is always, or almost always, a bad thing *about* that person—because, had she lived, she would have had more good experiences, hence her permanent $death_s$, in a certain sense, is a "deprivation" for a counterfactual counterpart of that person—it does *not* follow that the permanent $death_s$ of a higher-level or Kantian real person is ever a bad thing *for* that person. This is simply because $death_s$ has no personal subject for whom *anything* can ever be a good thing or a bad thing.

Moreover, not even Suits has recognized that although it is quite true that the permanent $death_s$ of a higher-level or Kantian real human person is never either a good

thing or a bad thing for the one who dies, simply because $death_s$ has no personal subject, nevertheless it does *not* follow that the $death_p$ of that very person is not also a good thing or a bad thing *for* that very person. By its very nature, $death_p$ has a living personal subject who is also in the process of dying; and, as I will argue, very often or even usually the $death_p$ of a higher-level or Kantian real human person is, tragically, a bad thing for that very person.

And this brings us to the fifth and final crucial ambiguity about the concept of death. This concerns the question of *whose death* is at issue, and in particular whether it is

(i) *my own* death, or
(ii) someone else's death,

that is at issue. The difference between my own death and the death of another higher-level or Kantian real human person is fundamental, whether we are thinking about $death_s$ or $death_p$. This, in turn, is because although we necessarily *have* first-person access to the contents of our own lives, we necessarily do *not* have first-person access to the contents of the lives of other higher-level or Kantian real human persons. Otherwise, we would *be* those other persons.

Differently put, "the problem of other minds" applies every bit as directly to the *deaths* of higher-level or Kantian real human persons as it applies to the *lives* of such persons. Necessarily, by the nature of my essentially embodied mind, I am both pre-reflectively consciously aware and also self-consciously directly aware of my own real personal life, but not of anyone else's real personal life. It follows that my own death, whether it is my $death_s$ or my $death_p$, *necessarily is no one else's death*. In this sense, we necessarily die alone, just as we necessarily live our lives alone. We are, to be sure, always living our lives alongside others' real human personal lives, and in more or less direct interaction and solidarity with others's real human personal lives. So in that sense, we always live our lives *with* other real persons' lives. But we do not *live* those lives, only our own. Mutatis mutandis, we are always dying alongside the deaths of other real human persons like us, in more or less direct interaction and solidarity with those others' deaths, and in that sense we are always dying *with* the deaths of others. But we do not *die* those deaths, only our own.

It is not my purpose in this chapter to engage directly with the problem of other minds in relation to the nature and value of death. For the record, I do hold that since higher-level or Kantian real human persons' minds are essentially embodied, and therefore are necessarily present throughout their living bodies, and made immediately manifest by the intentional orientations, movements, and expressions of those living bodies, and because our awareness of these manifestations-of-mind is non-conceptual and non-inferential, then we can directly grasp the mental lives of others to a greater or lesser extent, via what I call "empathic mirroring."[311] But this empathic mirroring presupposes the brute fact that the other higher-level or Kantian real person's life is necessarily external to my own life.

Empathy is *not* mind-melding or mind-transference. Correspondingly, we can empathically mirror the deaths_p of others, but necessarily we do not die the deaths_p of others. Someone else's death_p is necessarily never a death_p that is *for* the one who dies. As a consequence, necessarily the deaths_s of others are not facts having a direct and specific bearing on us. Someone else's death_s is necessarily never directly and specifically something that is *about* the real human person who dies.

That all having been said, there is still the deeply serious question: How ought we to think about the inevitable deaths of our loved ones? I have argued that someone's else's death, whether it is that person's death_s or her death_p, cannot be subjectively experienced by us. But we can and do, at second-hand, via empathic mirroring, subjectively experience the death_p of our loved ones, family, and close friends. And then, once they have died, we have post-life knowledge about their deaths_s. Morever, this post-life knowledge is always, other things being equal, an inherently bad thing for the one who survives: we *miss* them, so much, and intensely grieve for them. It is therefore not a sublime experience, on the contrary, it is a truly awful experience, intense suffering—although I do also think that one's grief *can*, in at least some contexts, be positively inflected by the thought that the other's death_p was an inherently good thing for her.

In any case, I think that the only thing that could possibly be worse than having post-life knowledge about the deaths_s of your loved ones, is never to have loved anyone or to have been loved by anyone. So the very same thing that makes a life lived with people you love an authentic joy, and as close to the full realization of deep happiness as it is possible to come in this thoroughly nonideal natural and social world, also makes knowing about their deaths_s a truly terrible experience. As far as I can see, there is simply nothing that can be done about this moral paradox, *except* to try with all your heart to live and die alongside your loved ones in such a way that nothing that really matters between you and them is ever left unexpressed, or undone. Even if you badly *fuck up* in wholeheartedly so trying. As one of the characters in Robert Bresson's amazing 1945 film, *Les Dames du Bois de Boulogne*, trenchantly observes:

> Il n'y a pas d'amour…. Il n'y a que des preuves d'amour. ("There is no love…. There are only proofs of love.")

Relatedly, I do also think that drama, films, literature, and other art forms such as voiced music and painting, especially those associated with religious traditions, are far better equipped to present the moral content and implications of these profoundly sad, brute facts about the rational human condition than philosophy is—with the sole possible exception, I think, of Plato's *Apology*, *Crito*, and *Phaedo*.[312] In any case, in this chapter I am not going to attempt it beyond what I have already said.

6.3 WHY DEATH$_S$ IS NOT LIVED THROUGH

Death$_s$, the state of being dead, is the same as the permanent cessation or end of a higher-level or Kantian real human person's *consciousness*. Why so? The answer is that, since consciousness is necessarily and completely neurobiologically embodied in a suitably complex living organismic system, the permanent cessation or end of organismic life in our own animal bodies necessarily entails the permanent cessation or end of our consciousness, which in turn necessarily entails the permanent cessation or end of our personal lives. In short, death$_s$ is a three-way *thanatological identity*:

> The end of my living animal body = the end of my essentially embodied consciousness = the end of my life.

To be sure, there are some real-world cases of full resuscitation after bodily processes have actually shut down, when the overall complex dynamic system of the human organism remains temporarily in a hiatus-state, capable of reactivation. And a temporary life after the death$_s$ of the body—say, a reanimation after the cryogenic preservation of one's corpse—is indeed not only conceptually or logically possible, and synthetically or really possible, but also nomologically possible. Furthermore, it may also *seem* that a sempiternally endless or infinite higher-level or Kantian real human person life after the actual death$_s$ of the human body, or even without any actual death$_s$ of the body—immortality—is conceptually or logically possible. But even so, that appearance of conceptual or logical possibility has nothing directly to do with the real metaphysics, epistemology, or ethics of higher-level or Kantian real human persons. Indeed, and sharply to the contrary, I will argue later that the appearance of the possibility of human immortality is nothing but a powerfully deceptive cognitive illusion: *the very idea of human immortality is incoherent and a priori impossible*. Then it a priori necessarily flows from the nature of a higher-level or Kantian real human person's life that she will die, and that her death$_s$ will be, just like her personal life and her death$_p$, once and forever:

> O, I die, Horatio;
> The potent poison quite o'er-crows my spirit:
> I cannot live to hear the news from England;
> But I do prophesy the election lights
> On Fortinbras: he has my dying voice;
> So tell him, with the occurrents, more and less,
> Which have solicited. The rest is silence.[313]

One striking consequence of this conception of death is that my own death$_s$ cannot be subjectively experienced by me. My own conscious, intentional, caring, rational human

life will permanently cease or end, but I will never subjectively experience the state of being dead. For $death_s$ has no personal subject. As Wittgenstein puts it in the two propositions from the *Tractatus* that are the fourth and fifth epigraphs of this chapter:

> [I]n death … the world does not change, but ceases…. Death is not an event of life. Death is not lived through.

In other words, our own conscious, intentional, caring, self-conscious, rational human minded animal lives will go on and on and on—until they simply end forever. Full stop, and "the rest is silence." Each such forever-silencing full stop on a conscious, intentional, caring, self-conscious, rational human minded animal life will be essentially unique, and thus my own $death_s$ will be essentially unique: it will be *my very own* $death_s$. In that sense, as I mentioned already, each of us necessarily dies alone.

This does not mean that other people and loved ones cannot be gathered around us as we finish the process of dying, or that things cannot positively or negatively affect us in an extrinsic or third-person sense after we die—both of these are really possible, and frequently actual. It means only that my $death_s$, just like my $death_p$, and just like my own life, is necessarily my very own. Out of the materials given me, over which I had little or no control, in a highly-structured, thoroughly nonideal world that I did not choose or create, *I freely shaped my life and my death, using just those materials and within just those constraints*. So it is my very own, no one else's, and nothing else's. Hence my $death_s$ has a "my very own-ness" in essentially the same sense that necessarily a single-authored book written by me is *my very own book*, even despite all the grateful acknowledgments to others who helped it come into existence, and even despite its readers, who can think about my book in ways over which I have no control, and who can keep my book alive (or kill it by critical abuse, or let it die by intellectual neglect) even when I am not.

In any case, my very own essentially unique $death_s$ cannot be subjectively experienced by me either as an intentional content or as an intentional object. For if my $death_s$ were either an intentional content or an intentional object of my subjective experience—as in so-called "after-death experiences"—then obviously that would imply the existence of my subjective experiences, and thus imply the existence of my own conscious, intentional, caring, rational human minded animal life. On the contrary, however, my permanent $death_s$ is my permanent annihilation and non-existence. Similarly a full stop, or period, is the end of a sentence and inherently belongs to the sentence as a proper part of its syntactic structure, but is not itself one of the words or phrases in the sentence. Punctuation formally or structurally terminates what is said by a sentence, but by itself does not say anything. So too, my very own essentially unique $death_s$ will be a termination that is cognitive, affective, practical, and vital *syntax*, but not cognitive, affective, practical, and vital *semantics*. Thus my very own $death_s$ will belong to the immanent structure of my conscious, intentional, caring, rational human minded animal life, but not to its vital stuffing, or occurrent mental

content and objects. Or in other words, my $death_s$ is nothing more and nothing less than *the immanent terminating form* of my very own life. It confers a definite, defining constraint and limit on the scope and shape of my entire life. It permanently *fixes* that very scope and that very shape. It makes my entire life *whatever it was for me*, once and forever.

These points all hold as much for one's own *natural death*, for example, one involving the gradual decline of my powers and well-being in old age, as they do for the various possible *strange deaths* arising from the well-known personal identity considerations and thought-experiments that I reconsider in *Deep Freedom and Real Persons*, chapter 7: real-world conjoined twin "fission" cases, fictional Star-Trek-like "transporter" cases, fictional "pseudo-Napoleon" cases, fictional "Lefty" and "Righty" fission cases (that is, simultaneous left-brain and right-brain transplants from real persons with split-brains/neocommissurotomy, into two new recipient bodies), and so-on.[314]

Thus when I am the split-brain case who is replaced by Lefty and Righty, *my very own* life ceases, full stop, and *their* lives begin. I do not subjectively experience my own $death_s$, because death is necessarily never subjectively experienced, yet it is a definite limit on my life just the same, at some definite time *t*. Lefty and Righty, both living at time $t + n$, each share all my memories, together with further and different present conscious experiences in different living animal bodies. But neither of them is *me*, because my very own life ended at *t* and thus all of my subjective experiences full-stopped right there. The spatiotemporal, neurobiological, and phenomenological structures of my very own life are all intrinsic to my real human personhood. Indeed, I am just my complete, finite, and unique life—so when it full-stops, necessarily I full-stop too. The natural or objective time of my $death_s$, necessarily occurring after my $death_{p,}$is the literal end of my personal or subjective time, the literal end of my "having the time of my life." So finite durations of time and my death, whether it is my $death_p$ or my $death_s$, are a priori necessarily connected.

In this way, one's own $death_s$ is nothing more and nothing less than an immanent structural and inherently temporal terminating constraint and limit on the occurrent mental content of one's own complete, finite, and unique rational human minded animal life, again like punctuation at the end of a sentence, but also with a metaphysical time-stamp that completes and rounds off all the events of a single personal life. As they say, time is of the essence.

Failing to recognize this, however, we naively imaginatively project a first-person standpoint on $death_s$ from the *other* side of this time-stamped limit, thereby generating the strong impression that $death_s$ somehow strangely belongs to the content of our lives, like a ghostly afterword or postscript. But this is a psychological illusion with serious existential and moral implications.

6.4 THE INCOHERENCE AND IMPOSSIBILITY OF PERSONAL IMMORTALITY

This strong but philosophically naïve impression that $death_s$ is a ghostly "event of life" that is, or anyhow can be, "lived through," in turn, gives rise to the even more serious conceptual illusion that higher-level or Kantian real human personal *immortality* is a coherent notion. But in fact, we do not have the slightest idea how the concept of sempiternally endless temporal extension or infinity applies to the concept of the life of a higher-level or Kantian real human person, far less to the concept of the life of any other sort of real human person. So it also turns out that immortality is a priori impossible for creatures like us.

I will establish these points by briefly unpacking and then criticizing two of the most influential and important discussions of the nature and value of immortality, Bernard Williams's "The Makropulos Case: Reflections on the Tedium of Immortality"[315] and John Martin Fischer's reply to Williams, "Why Immortality is Not So Bad."[316]

(1) Williams on the Tedium of Immortality

In his justly-famous paper, Williams wants to argue for two theses:

> (i) other things being equal, death is a bad thing for the person who dies, and
> (ii) immortality would be, where conceivable at all, intolerable.

The argument for thesis (i) has three steps.

First, there are certain desires, that Williams calls "categorical desires," which are desires that are unconditional with respect to rational human life, in that we want them to be satisfied whether or not we are alive to experience them. For example, rational suicide, understood as the reasonable desire to be $dead_s$, is such that the rationally suicidal subject wants this desire to be satisfied even though he will not be alive to experience that state. Although Williams does not use this term specifically, let us call any similar inherently $death_s$-related or rationally suicidal desire—for example, the desire that event *X* happens *N* days after one's own suicide—a *negative categorical desire*.

Second, correspondingly, a conscious rational subject can categorically desire things in a *positive* way, beyond his own death. For example, I could intensely desire to be the Nobel Prize winner for Literature in 2057, exactly 100 years after Albert Camus won his prize in 1957, the year of my own birth, and, assuming that no other philosopher wins it in the meantime, thereby become the first *philosopher* to win the Prize since Jean-Paul Sartre in 1964—even though the likelihood of my actually living beyond my 80s is fairly small. More generally, positive categorical desires can include the desire to go on living after one's own actual death, so that many future desires will come into existence and be satisfied.

Third, therefore, as long as the conscious rational subject has positive categorical desires, then it is a bad thing for that person to die.

Importantly, according to Williams, categorical desires are inherently contingent in that we do not *have* to have them, or at least we do not *always* have to have them. Indeed, on the supposition that as a matter of contingent fact someone has no positive categorical desires, or that any positive categorical desires that person previously had have now been extinguished, then death_s could be a good thing, and one could have a good reason to die, in that it satisfies a negative categorical desire. For example, Elina Makropulos, the fictional protagonist of *The Makropulos Case*, has been granted immortality, but within the first three centuries of her sempiternally endless or infinite life, starting at age 42, she has also lost all her positive categorical desires. So then, at age 342, she negatively categorically desires to be dead_s, and therefore has a good reason to die.

This provides the conceptual segue to Williams's argument for his thesis (ii), which has four steps.

First, it is a necessary condition of my being immortal that the very same person—namely, I myself, as I am now, with a certain set of memories, and a certain character—goes on living, and does not change identities over time. The idea that I myself am continually being reborn as a new person, as opposed to merely being reincarnated in a new body, is incoherent.

Second, as time passes, all of the experiences it would be possible for me to have, are eventually had. Then after that time, necessarily, a state of boredom, indifference, and coldness—in Williams's nice phrase, "joylessness"—sets in. Presumably, joylessness consists in having *no* desires that must be satisfied

> either (i) as actually experienced by me with joy, hence conditional on my being alive to experience them (joyful-life-conditional desires),
> or (ii) as would be experienced by me with joy, if, contrary to highly probable fact, I continued to live (positive categorical desires per se).

Third, for this reason, living forever would be infinitely joyless, and, in particular, infinitely boring.

Fourth, therefore immortality would be intolerable.

(2) Fischer on How Immortality Could Be a Good Thing

According to Fischer, by virtue of his argument for the intolerability of immortality, Williams's account negatively implies two necessary conditions on the tolerability of immortality:

> (i) *the identity condition*, which says that the person who lives on must remain the same person over time, and

> (ii) *the attractiveness condition*, which says that the person's future life must be appealing, that is, not filled with pain and/or suffering, and not joyless—in particular, and perhaps most importantly, not boring.

In view of those necessary conditions, Fischer then claims that Williams's argument makes three questionable assumptions, and fails to recognize one crucial distinction, hence it is an unsound argument.

The first questionable assumption that Williams makes is that in order for immortality to preserve identity over time, future activities cannot be completely absorbing, since then the subject would lose herself, and therefore *her self*, in them, and could not preserve her identity over time. But as Fischer correctly points out, it is one thing for the *content* of an experience to be completely absorbing, and quite another for an experience to be *unowned* by a distinctive, synchronically and diachronically identical self. More generally, completely absorbing experiences in the content-sense can also be owned by the very same self at any given time and over time.

Williams's second questionable assumption is that in order for immortality to be attractive, it must consist in one single activity that in turn would eventually become joyless and boring. But on the contrary, Fischer plausibly argues, immortality could consist in a plurality of activities, and it is not at all clear that this plurality would itself ever be joy-exhaustible or become boring in the way that a single activity could.

And Williams's third questionable assumption is that in order for immortality to be attractive, all experiences in the subject's future sempiternally endless or infinite life have to be pleasurable, even though they all would eventually become joyless and boring. But on the contrary, according to Fischer, since *finite* or terminating lives can be overall very good even if there is a certain amount of pain/suffering, joylessness, and boredom in them, then there is no good reason to think that a sempiternally infinite or endless life could not be similarly composed.

In addition to these three questionable assumptions, according to Fischer, Williams fails to recognize a crucial distinction between

> (i) *self-exhausting pleasures*, which aesthetically and/or hedonically terminate themselves and are inherently non-renewable for the subject, either (ia) because they turn out, in the event, to be disappointing (for example, the prospectively amazing New Year's party that is not so very amazing after all, indeed quite the contrary) or (ib) because they are complete in themselves (for example, the intense thrill of climbing Mount Everest, that one never needs or wants to repeat, having "been-there, done-that"), and
> (2) *repeatable pleasures*, that do not exhaust themselves and are inherently worth experiencing again and again.

Self-exhausting and repeatable pleasures can, to some important extent, be relativized to individuals and contexts: what counts as self-exhausting or repeatable for one individual

or in one context, need not count as self-exhausting or repeatable for another individual or in another context. Moreover, repeatable pleasures should not, in general, be obsessively or mechanically repeated, but instead require appropriate distribution or patterning over time. Now Williams seems to assume that all pleasures will ultimately be self-exhausting in the condition of immortality. But, on the contrary says Fischer, there is no good reason to believe that there cannot be endlessly or infinitely repeatable pleasures in a sempiternally endless or infinite life, provided that these pleasures are appropriately distibuted or patterned over time.

So, taking Williams's three questionable assumptions together with his failure to recognize the category of repeatable pleasures, his conclusion does not follow. On the contrary, Fischer concludes, immortality *could* be a good thing.

(3) Some Worries About Williams's Account and Fischer's Account Alike

For the purposes of my criticism of Williams and Fischer alike, by "a finite or terminating rational conscious human life" I will mean *a higher-level or Kantian real human personal life, with permanent death*$_s$ *at the end of it*. Then, correspondingly, by "immortality" I will mean *a sempiternally endless or infinite higher-level or Kantian real human personal life*.

Granting that, then we need to distinguish between

> (i) a finite or terminating higher-level or Kantian real human personal life that is relatively short, say, lasting up 120 years in duration as an absolute maximum, but no longer than that,
> (ii) a finite or terminating higher-level or Kantian real human personal life that is super-long, say, any finite number of years greater than 120 in duration, including of course Elina Makropulos's 342 years, and
> (iii) a higher-level or Kantian real human personal life that is sempiternally endless or infinite.

The deep issue raised by this threefold distinction is how precisely we are to understand the concept of *endlessness* or *infinity* when it is applied to the concept of a higher-level or Kantian real human personal life.

Now a real human personal life like ours, simply by virtue of its being human and therefore having a necessary connection with organismic life, occurs in rather limited portions of space, and also has a certain temporally definite biological sequencing related to growth, maturation, aging, eating, sleeping, breathing, blood circulation, heart activity, neuronal activity, hormonal activity, ranges of body temperature, and so-on. In other words, a higher-level or Kantian real human personal life is inherently filled with spatial and biotemporal parameters of various kinds. By sharp contrast, the only well-defined concept of endlessness or infinity we have is fundamentally mathematical, and here there is an important distinction between

> (i) *denumerable* infinities, involving one-to-one correspondence with the set of natural numbers/positive integers), and
> (ii) *non-denumerable infinities*, which systematically outrun one-to-one correspondence with the natural numbers/positive integers, for example, the power set of the set of natural numbers.

We can meaningfully add this dual mathematical concept of endlessness or infinity to the concept of sempiternally successive time, and then understand the idea of a sempiternal endlessness or infinity that is either denumerable or non-denumerable. But, supposing that we do have some conceptually competent grasp of the temporal-mathematical concept of sempiternal endlessness or infinity, nevertheless I do not think we have the slightest idea of *how this concept meaningfully applies to the concept of a higher-level or Kantian real human personal life*, given the necessary connection between such a life and an inherently spatially-limited and temporally definite biologically-sequenced organismic life of a specifically human sort.

For example, in an endless or infinite amount of time, since every denumerably infinite series has the same cardinality, the very same higher-level or Kantian real human person could visit every single point in any denumerably infinite space. And even though, necessarily, every higher-level or Kantian real human person, by virtue of their specifically human organismic lives, grows, matures, and ages throughout those lives, that very same person could also somehow exist for an endlessly or infinitely long time without growing, maturing, or aging, like Elina Makropolus. But none of this makes any sense. How could the constitutive moments of a single higher-level or Kantian real human person's life map one-to-one to *all* the points of any infinite space? Does Elina Makropulos need to eat, or not? If so, what are her digestive processes like? Does she need to sleep, and if so, why? Is she constantly exchanging heat, energy, and matter with the environment, like every other compex dynamic system that is an animal? Is she subject to entropy? And so-on. Hence I do not think we have the slightest idea of what the concept of "higher-level or Kantian real human personal immortality" really means.

Correspondingly, on the charitable assumption that they are *actually* making sense, I think that Williams and Fischer are *actually* talking about a finite or terminating higher-level or Kantian real human personal life *that is super-long*, and not about higher-level or Kantian real human personal *immortality*, which is in fact an incoherent notion.

On the one hand, then, Williams is absolutely right that there is something deeply questionable about the very idea of immortality for creatures like us; but also Williams is quite wrong that a finite or terminating higher-level or Kantian real human personal life that is super-long would be intolerable, for all the reasons that Fischer gives. And on the other hand, Fischer is absolutely right that Williams's argument for the intolerability of immortality is unsound; but also Fischer is quite wrong that he has shown *anything* about

how higher-level or Kantian real human personal immortality could be good, since the very idea of such a thing is incoherent.

In fact, immortality for creatures like us, higher-level or Kantian real human persons, is a priori impossible because its very idea is incoherent, and more precisely because its possibility is ruled out a priori by the very idea of a higher-level or Kantian real human personal life. Our conscious, intentional, caring lives as higher-level or Kantian real human persons are finite but unbounded, like the surface of a sphere. Or to make the same point slightly differently, since every such conscious, intentional, caring life necessarily has egocentric centering, it is like the shape of the visual field, which is the interior of a finite sphere projected perspectivally outwards from a single oriented region on that interior surface. Our subjective experience of the finite unboundedness of the interior of this orientable, thermodynamically irreversible, egocentrically centered, complete, unique perspectivally-projected life-sphere—a sphere that is completely filled with intentional contents, intentional objects, and ourselves, fully embedded in a thoroughly nonideal natural and social world and along with other conscious subjects and other living organisms—is as close to immortality as we will ever get because it is as close to immortality as it is a priori possible for creatures like us to get. Since, according to The Minded Animalism Theory of personal identity,[317] every real human person is literally identical to each and all parts of her own complete, finite, and unique essentially embodied life-process, and since each real human person's life-process thereby has both a definite unique beginning and also a definite unique ending, then the very notion of a "sempiternally endless or infinite life" for creatures like us is a priori impossible.

Furthermore, and perhaps most poignantly, to hope for immortality, or to desire and long for immortality, is a tragic conceptual and metaphysical mistake, a serious cognitive illusion. As the Tractarian Wittgenstein clearly saw—his attention having been duly concentrated by the horrors of front line action on the Eastern Front in the Great War—this hope, desiring, or longing for a sempiternally endless or infinite life in effect just endlessly or infinitely puts off till tomorrow what you can, really necessarily, only ever feel, choose, or do right here and right now, today, over and over and over again, until, inevitably, you die. "He lives eternally who lives in the present." To hope, desire, or long for immortality is therefore a fundamental denial of your own innately-specified capacity for principled authenticity, and in this way it constitutes a special form of nihilism that Simon Critchley aptly calls *passive* nihilism.[318]

So here is where Existential Kantian Ethics and early Wittgenstein's *Tractatus* meet up with Nietzsche's later philosophy. Indeed, in my opinion, Wittgenstein's thought about living eternally in the present is essentially the same as the one Nietzsche had about "the greatest stress" and eternal recurrence. Both of these thoughts express a profound dual insight about the nature of principled authenticity and about the self-undermining passive nihilism that constantly tempts us in the form of the seemingly benign and natural desire for endless or infinite life.

6.5 DEATH_S IS NEITHER A BAD THING NOR A GOOD THING FOR THE ONE WHO DIES—YET DEATH_P CAN BE EITHER A GOOD THING OR A BAD THING FOR THE ONE WHO DIES

Because death_s has no personal subject and therefore no first-person standpoint, death_s is *nothing at all* for the real human person who dies, hence death_s *is neither a good thing nor a bad thing for the real human person who dies*. To put this point in quasi-Epicurean terms, "where we are, death_s is not, and where death_s is, we are not." Death_s can be a good or bad thing *about* the real human person who dies, but never *for* the real human person who dies. In other words, it is true that death_s always involves various good or bad external or extrinsic facts about the real human person who dies, from a third-person standpoint—for example, that it would have been a good thing had she lived longer; or that certain bad things about that person come to light after she dies, etc.—but none of these are good or bad internal or intrinsic facts *for* the real human person who dies, from her first-person standpoint. Nevertheless, at the same time, since necessarily death_p is *infra*-life and not post-life, it is also really possible for someone's process of dying to be either a good thing or a bad thing for the real human person who dies.

Using the same dialectical strategy as section 6.3, I will establish these points by critically analyzing two of the most important and influential discussions of the nature and value of death, Thomas Nagel's justly famous "Death,"[319] and David Suits's important critical reply to Nagel, "Why Death is Not Bad for the One Who Died."[320]

(1) Nagel On the Nature of Death

The core philosophical question raised by Nagel's essay is this:

> If we assume that death is the unequivocal and permanent end of our personal existence, so that any question about immortality is ruled out for the purposes of argument, then is death a bad thing for the one who dies?

And Nagel's strong concluding answer to his own question is that *yes*, under the assumption that the life of a human person is finite and terminating, then death is *always* a bad thing for the one who dies. In turn, Nagel's argument for his strong conclusion has ten basic steps.

First, if death is bad, this is solely because of what it deprives us of, not because of any positive features it has, unlike life.

Second, what is fundamentally good about life are certain states, conditions, or types of activity: being alive, doing certain things, and having certain experiences. We could call this *having-a-life* or *being-the-subject-of- a-life*. So what is fundamentally good about life is having-a-life or being-the-subject-of-a-life. Even if the contents of a life are bad, perhaps

very bad, the very fact of having-a-life or being-the-subject-of-a-life is intrinsically good.[321]

Third, these two theses imply a distinction between

(i) the essentially positive character of the goodness of life, and
(ii) the essentially negative character of the badness of death.

Fourth, as to the essentially positive character of the goodness of life, we can say

(i) that life has various benefits whether intrinsic, instrumental, or otherwise relational,
(ii) that the value of life does not attach to mere organic survival, since mere organic life in a coma is valueless, and
(iii) the goods of life can be multiplied by time—although not necessarily continuously over time, since suspended animation or cryogenic preservation of the body, together with reanimation, seems perfectly consistent with the multiplication of goods during the reanimated period—so more of the goods of life is better than less of those goods.

Fifth, as to the essentially negative character of the badness of death, we can say that death is an evil because it consists in *the deprivation or loss of life*, rather than in *the state of being dead*. For personal nonexistence, as such, is not necessarily a bad thing. The temporary suspension of life is entails no disvalue, so long as it does not reduce the total lifespan; and most of us are not bothered by the fact that we did not exist before we were born.

Sixth, corresponding to the first five points, here are three hard questions about the badness of death:

(i) how can anything be bad for someone without its also being an unpleasant experience for her?,
(ii) the state of being dead is without a subject or first person to experience it, so how could it ever be bad for anyone?, and
(iii) how can death be bad if pre-natal nonexistence is not a misfortune (Lucretius's question)?

Seventh, here is the answer to question (i). Many goods and bads for persons are not directly attributable to the intrinsic character of their momentary or durational states of mind but instead to their entire life-histories, including various diachronic extrinsic relations to their earlier and later selves, as well as both diachronic and synchronic extrinsic relations to other persons, events, and things. So, since many goods and bads are extrinsic relational and not (merely) intrinsic features of persons, then someone can suffer misfortune in a purely extrinsic relational sense even when he is not in a position to

recognize that misfortune, and experience it as unpleasant—for example, when ignorant, asleep, fainting, unconscious, in a coma, non-rational, or dead.

Eighth, here is the answer to problem (ii). The subject of death is the human individual, the subject of a single life, which may or may not include his or her personhood but necessarily includes his or her personhood if he or she has ever actually been a person. And this person is someone who can suffer extrinsic relational harms even if she is not in a position to experience those harms as unpleasant. For example, someone could suffer all sorts of extrinsic relational miseries (betrayal, the death of his loved ones, theft of his property, slanderous damage to his good name and reputation, the loss of his rational faculties through disease or injury, etc.) without experiencing these as harms as unpleasant. Hence the very same subject can also suffer these extrinsic relational harms after death.

Ninth, here is the answer to question (iii). The time before a human individual's actual birth is not a time when that individual could have been alive, because a human individual's actual beginning or birth is a necessary condition of his or her individuality, so nothing can really matter to the human individual until after birth. Therefore only post-natal nonexistence can count as the death of the individual and be a misfortune to the individual.

Tenth and finally, the badness of death consists in the non-realization (or deprivation) of future possibilities of having-a-life or being-the-subject-of-a-life, including both intrinsic goods and extrinsic relational goods. Hence the earlier one dies the worse it is, and the later one dies the better it is. Even despite the fact that our lives have natural limits, since one's own life appears from the first-person standpoint to be essentially open-ended and unlimited, and provided that there is no limit to the amount of life it would be good to have, then death is always a bad thing for the one who dies.

(2) Suits Against the "Deprivation" Account of the Badness of Death

In his critical reply to Nagel, Suits argues that a dead person can neither know, appreciate, or in any possible a way experience any effects of death. As Suits puts it, "death is a singularity for each of us."[322] Thus death is *the terminal limit of a life*, not a part of a life. But the only way a person can be harmed is by actually suffering pain (primitive intrinsic harm) or prospectively suffering pain (derivative intrinsic harm) of some sort. Therefore a person cannot be harmed by death. So Lucretius was correct when he said that "death is nothing to us."

Moreover, death is not a deprivation on any reasonable understanding of what deprivation is. Deprivation is failing to get some good things that were in some sense expected, and then knowing, appreciating, or somehow experiencing the failure to get these things. But the dead person never feels deprived either primitively or derivatively, precisely because she never feels anything at all.

If we have interests and they are defeated or frustrated, then we suffer pain and are harmed. But death is not the defeat or frustration of our interests: it is merely the permanent disappearance or permanent *vacating* of our interests. And if we do not have any interests, then they cannot be defeated or frustrated. Hence we cannot be harmed by the permanent

vacating of interests caused by death. And therefore the deprivation account does not show that death is bad in any recognizable sense for the deceased.

While Nagel's deprivation account relies on an actual-life vs. counterfactually-longer-life comparison, this comparison does not entail that death can be bad for the person who died, because the person who dies has an *actual* life, not a *counterfactua*l life. Counterfactual comparisons can show something *about* someone who dies, but they are nothing *for* the person who dies. Thus death is never anything for the person who dies, either a bad thing or a good thing. And as a consequence, death is never a bad thing for the one who dies.

(3) Some Critical Worries About Nagel's Account and Suits's Account Alike

As we have seen, Nagel's account says that death is *always* a bad thing for the person who dies, whereas Suits's account says that death is *never* a bad thing for the person who dies. The fundamental problem with both views is that neither Nagel nor Suits distinguishes carefully between

(i) the state of being dead, death_s, and
(ii) the process of dying, death_p.

On the one hand, then, I think that Suits is absolutely correct about death_s. Since death_s has no subject or first-person, then death_s is neither a good thing nor a bad thing for the real human person who dies. Hence death_s is never a bad thing for the person who dies. So Nagel is wrong about death_s.

But on the other hand, when we consider death_p, things come out somewhat differently.

First, it is true that sometimes more life will inevitably lead to person-destroying suffering, for example, degenerative diseases like Alzheimer's. Similarly, sometimes more life will inevitably lead to some irremediably or irreparably monstrous evil or evils being freely committed by that person, for example, Dostoevsky's Raskolnikov in *Crime and Punishment* shortly before he axe-murders the old lady pawnbroker and her sister. And again, sometimes more life will inevitably lead to the irremediable or irreparable self-destruction of someone's own integrity, for example, an otherwise decent person shortly before he freely succumbs to some terrible temptation—say, knowingly and without being forced, allowing an innocent person to be tortured to death by others, simply in order to move ahead in the Nazi command-hierarchy, or simply in order to receive some sum of money by a Mafia payoff, etc.—and irrevocably compromises himself. Then in all such cases, an earlier death_p would be a good thing for the person who dies. So Nagel is wrong that death_p is always a bad thing for the person who dies. On the contrary, *death_p is sometimes a good thing for the person who dies*.

Second, for the purposes of our argument, we can suppose that it is true, as Existential Kantian Ethics holds, that we are morally obligated to pursue principled authenticity. And we can also suppose further that an authentic principled life is necessarily a finite or

terminating life with an internal narrative structure and closure. Then if you have failed to achieve or realize principled authenticity, at least partially and to some degree, by the time you die, then $death_p$ is a bad thing for the real human person who dies. So Suits is wrong that $death_p$ is never a bad thing for the one who dies. On the contrary, *$death_p$* is *sometimes, and indeed all-too-frequently, a bad thing for the higher-level or Kantian real human person who dies*.

6.6 UNTIMELY $DEATHS_P$ AND WHY WE SHOULD RAGE AGAINST THE DYING OF THE LIGHT

In view of what I have already argued, here are three theses about the morality of one's own death.

First, the concept of an *untimely $death_p$* is fully meaningful and also has actual instances.

Second, an untimely $death_p$ is necessarily an inherently bad thing for the higher-level or Kantian real human person who dies in this way, regardless of the other ways in which it might also be bad—for example, in an intrinsic or first-person way, by way of its bodily painfulness, or, in an extrinsic relational or third-person way, by way of its being contrary to the person's self-interest, or its having bad consequences for others. This is precisely because, in the process of dying, that higher-level or Kantian real human person fails to achieve or realize principled authenticity, at least partially or to some degree. Hence by dying in this way she has, tragically, to that extent, *wasted her life*.

Third and corresponding to the other two theses, I also want to defend a thesis I will call

> Death's Excluded Middle:
>
> All $deaths_p$ of higher-level or Kantian rational human minded animals are either *untimely*, in that they are inherently bad for the higher-level or Kantian real human person who dies, or else they are *timely*, in that they are inherently good for the higher-level or Kantian real human person who dies, and they are never both untimely and timely.

In other words, necessarily there are no $deaths_p$ that are neutral or null with respect to intrinsic moral value, understood in terms of principled authenticity. This is precisely because the subject of $death_p$ is necessarily always a higher-level or Kantian real human person, and such creatures are necessarily never neutral or null with respect to intrinsic moral value, understood in terms of principled authenticity.

From these three theses, then, it follows immediately that

> either (i) all deaths_p are untimely and thus inherently bad for the higher-level or Kantian real human person who dies,
> or else (ii) only some deaths_p are untimely, because some other deaths_p are, on the contrary, timely deaths and thus an inherently good thing for the higher-level or Kantian real human person who dies.

As I have indirectly indicated already, my view is that (i) is false and (ii) is true. Hence we should all be endeavoring with all our hearts, throughout our lives, to have timely deaths_p. That is the core thought of The Rage-Against-the-Dying-of-the-Light Theory of the nature and moral value of one's own death.

And here is more of the rationale behind that core thought.

The process of a real human person's life is identically the same as the process of her dying. Hence a real human person's life is also her own death_p. Now the ultimate meaning or purpose of a higher-level or Kantian real human personal life is to achieve or realize principled authenticity, at least partially or to some degree. Therefore, to the extent that one *fails* to achieve or realize principled authenticity, at least partially or to some degree, then death_p is a bad thing for the person who dies. Many people's natural deaths_p are untimely in the sense that they occur in lives that do not exemplify principled authenticity at all. But this is not necessary, it is merely widespread. For not every natural death_p is an untimely one. One's own natural death_p, that is, one's own life up to the very moment of the beginning of the permanent condition of one's own death_s, *can* exemplify principled authenticity, at least partially or to some degree. Thus we have no sufficient reason to fear an untimely natural death_p, because as long as we are wholeheartedly trying to achieve or realize principled authenticity, then in fact we are *already* achieving or realizing principled authenticity, at least partially or to some degree. Then we are already *on the way*, already *embarked*, on the achievement or realization of principled authenticity. And as long as you are alive, sentient, and sapient, you can always change your life. So as long as you are alive, sentient, and sapient, *then there is always enough time left for everything that really matters*. Therefore, you ought to "rage, rage against the dying of the light."

This is shown by the very obvious fact that someone can have-a-life and be-the-subject-of-a-life, yet fail ever to choose or do anything meaningful or that achieves or realizes principled authenticity, at least partially or to some degree, and either just *drift listlessly* towards death_s or (what is perhaps even worse) *busily busy-bee* towards death_s. —Always making more and more money, more and more honey, always embodying the Spirit of the Hive, always being the good little capitalist boss, professional, or worker do-bee of the modern neoliberal democratic state.

In other words, it is really possible *to waste your life*. And that is a *tragedy*, in the specifically modern sense of that classical Greek and Aristotelian notion, which typically involves the actuality or real possibility of greatness of character in a certain higher-level or Kantian real human person, a correspondingly great character flaw in that real person

like us, a terrible downfall for that real person as a direct result of that great character flaw, and some sort of cathartic experience for the witnesses of this downfall, and so-on. Perhaps the most vivid literary expression of this is Shakespeare's Hamlet:

> Who would fardels bear,
> To grunt and sweat under a weary life,
> But that the dread of something after death,
> The undiscovered country, from whose bourn
> No traveller returns, puzzles the will,
> And makes us rather bear those ills we have
> Than fly to others that we know not of?
> Thus conscience does make cowards of us all,
> And thus the native hue of resolution
> Is sicklied o'er with the pale cast of thought,
> And enterprises of great pitch and moment
> With this regard their currents turn awry
> And lose the name of action.[323]

As Hamlet's fictional case shows, it is possible, tragically, to lack all purity of heart, lack all wholeheartedness, and lack all single-mindedness, and yet also to be fully self-conscious of this very lack. Hamlet is the ultra-self-conscious Prince of Denmark, the ultra-self-conscious Prince of Double-Mindedness, and the ultra-self-conscious Prince of Losing Heart alike. It is self-evident that Hamlet's sort of life and Hamlet's sort of death_{p} are both inherently bad and tragic, not inherently good. Thus it is self-evident, by practical negation as it were, that what I will call a *Contra-Hamlet*'s sort of life and a Contra-Hamlet's sort of death_{p} would both be inherently good and sublime, not inherently bad and tragic. The life of a Contra-Hamlet is a life in which principled authenticity is achieved or realized, at least partially or to some degree. Correspondingly, as I pointed out in *Deep Freedom and Real Persons*,[324] in order to make the very idea of a life of principled authenticity more concrete, we can think here of Socrates as represented by Plato in the *Dialogues*; of the heroically absurd "Knight of the Sorrowful Countenance," Don Quixote, in Cervantes's *Don Quixote*; of Kierkegaard's "Knight of Faith" in *Fear and Trembling*; of the "Idiot" Prince Myshkin in Dostoevsky's *The Idiot*; of Renée Falconetti's brilliant portrayal of Joan of Arc in Carl Theodor Dreyer's *Passion of Joan of Arc*; of Takashi Shimura's equally brilliant portrayal of the dying civil servant Kanji Watanabe in Kurosawa's *Ikiru*; and also of the real-life, therefore "human, all too human," and thus "sinner-saints," but still genuine moral heroes Abraham Lincoln, Mahatma Ghandi, Martin Luther King Jr., and Mother Teresa. And there are *many, many* unsung others just like them.[325] In my opinion, all of these Contra-Hamlets and sinner-saints, whether fictional or real-life, died deaths_{p} that were inherently good, sublime, and timely, just as they lived. So we should all be trying

with all our hearts to live and die like them, in our own unique contexts and in our own unique ways, in the time remaining to us.

So I will argue that *not all* deaths_p are untimely, and that *at least some* deaths_p are timely and therefore an inherently good thing for the higher-level or Kantian real human person who dies. So too, I will argue that a higher-level or Kantian real human person's death_p D is timely if and only if

> either (i) D is an inherently good thing for the higher-level or Kantian real human person who dies, because continued life would be in some way personhood-destroying for her,
> or (ii) D is not only an inherently good thing for the higher-level or Kantian real human person who dies, but also a supremely good thing for her, because by means of her process of dying she achieves or realizes principled authenticity, at least partially or to some degree.

In cases that fall under (i), death is an intrinsically good thing for the higher-level or Kantian rational human minded animal who dies, precisely because her dignity as a real human person is thereby preserved in the face of the real threat of its loss or irrevocable degradation. In cases that fall under (ii), death is also the highest inherently good thing for the higher-level or Kantian real human person who dies_p, precisely because her ultimate end or purpose as a real human person with dignity is thereby achieved or realized, at least partially or to some degree. This, in turn, is because principled authenticity is the Highest or Supreme Good for every higher-level or Kantian real human person.

The basic idea behind The Rage-Against-the-Dying-of-the-Light Theory of the nature and moral value of death, then, is this. Although $death_s$, the state of being dead, *is nothing for us*, nevertheless $death_p$, the process of dying, *is of overriding importance for us*. The ultimate significance of one's own death_p is contained necessarily and completely immanently within the essentially embodied, intentionally active, life-process of the higher-level or Kantian real human person whose death_s provides a unique, permanent closure on her entire life-process. Thus the ultimate significance of one's own death_p lies entirely and exclusively in what one actually chooses and does with one's own higher-level or Kantian real human personal life. The ultimate meaning of one's own life, which is identical to one's own process of dying, in turn, is just the global pattern or shape of the total set of specific diachronic and synchronic profiles of her higher-level or Kantian real human life-process and death-process—a global pattern or shape that is dynamically emergent from her active pursuit of principled authenticity, within the necessarily finite limits of the complete, unique, permanent, full-stop, time-stamped structural closure provided by her own death_s. And the rest, really and truly, is nothing but silence.

Hamlet's central and tragic, passively nihilistic mistake lay precisely in his thinking that there *could be* something for him after death_p, some sort of ghostly tag-end of his life viewed from the non-existent standpoint of his death_s. Otherwise he would not have put off endlessly till tomorrow what he could only ever have done eternally in the present. Single-

mindedness, purity of heart, or wholeheartedness—in a word, authenticity—is living as if Nietzschean eternal recurrence were true, and as if everything always really mattered right here and now. On the contrary, however, whether ultra-self-conscious, only ordinarily self-conscious, or even mostly un-self-conscious, Hamletian double-mindedness, impurity of heart, half-heartedness, or lack of heart—in a word, inauthenticity, or passive nihilism—*is living as if Nietzschean eternal recurrence were impossible, as if there could somehow be something more than a finite, unbounded life and death$_p$, something after death$_p$, the ghostly realm of death$_s$, an "undiscovered country, from whose bourne no traveller returns."* This is also to live as if nothing ever really mattered right here and now because you yourself are, for example, nothing but a fleshy deterministic or indeterministic and indestructible Turing machine eternally programmed for endlessly yielding the same result—presumably, '42'[326]—in a spaceless and timeless After-Life created and ruled by an infinitely distant all-powerful, all-knowing, and all-good God, The Divine Commander. So there is a set of very deep-running connections, essential analogies, and thus *elective affinities* between

(i) the belief in immortality,
(ii) existential inauthenticity,
(iii) passive nihilism,
(iv) the belief in Universal Natural Determinism and/or Natural Mechanism,
(v) the belief in theism combined with Divine Command Ethics, and
(vi) mindless obedience to the inherently rationally unjustified authority of the State and other State-like institutions.

But here I am verging on fundamental issues in what I call *political theology*, that I discuss in detail in *Kant, Agnosticism, and Anarchism* and in "Exiting the State and Debunking the State of Nature" (THE RATIONAL HUMAN CONDITION Vol. 1, essay 2.1).

6.7 The Morality of Euthanasia

Euthanasia is when a higher-level or Kantian real human person kills another such real person intentionally, and solely from the motive of mercy. A good example, taken from the movies, is when the Clint Eastwood character in *Million Dollar Baby*,[327] a boxing trainer, kills the Hilary Swank character, his permanently paralyzed boxing protegée. The specific motive of mercy-killing on the part of a merciful person *A* entails *A*'s sincere belief that the dignity of another real person *B* is being violated by continued life, together with *A*'s sincere belief that she can prevent or reduce this violation of *B*'s dignity, together with *A*'s sincere belief that killing *B* is the *only* way of stopping or preventing this violation of *B*'s dignity. Hence euthanasia is intentionally chosen and done solely in order to prevent

or reduce a dignity-violating harm to the real human person who is mercy-killed, and *not* for the good of the higher-level or Kantian real human person who mercy-kills.

Under what conditions, if any, is euthanasia morally impermissible, permissible, or obligatory?

An apparently basic distinction in this sub-region of The Web of Mortality is between

> (i) *active euthanasia*, which is mercy-killing someone by directly or indirectly intentionally causally intervening in that real human person's vital processes, and
> (ii) *passive euthanasia*, which is mercy-killing someone by intentionally not causally intervening in that real human person's vital processes.

But it has been plausibly argued by many moral philosophers, for example, by James Rachels, that there is no morally important difference between active and passive euthanasia.[328] If and whenever it is morally impermissible, permissible, or obligatory to mercy-kill by *intentionally intervening* in someone's vital processes, then it is also morally impermissible, permissible, or obligatory to mercy-kill by *intentionally not intervening* in that real person's vital processes, and conversely.

This "no-morally-important-difference" thesis can also be smoothly confirmed by looking at any actual or possible case of passive euthanasia that is deemed to be morally permissible or impermissible, and then slightly re-conceiving the case. This minor re-conception requires only that in a relevantly nearby possible world, the intentional act of killing by non-intervention also accidentally triggers a causal process which, by a deviant causal chain, also ends up causally overdetermining the killing by intervening directly or indirectly in the vital processes of the real person who is mercy-killed. So, for example, the compassionate doctor who is going to let his suffering patient die painlessly, in the very act of refraining from direct or indirect intervention in the vital processes of her patient, accidentally also triggers the injection of a drug that painlessly kills the patient at the very same moment she would have painlessly died by non-intervention. The presence of the overdetermining causal intervention obviously does not affect the existing moral permissibility or impermissibility of the doctor's act of passive euthanasia. And moral obligatoriness obviously requires moral permissibility. Hence there is no morally important difference between passive and active euthanasia, and therefore the active euthanasia vs. passive euthanasia distinction is not a basic moral distinction in this area.

It should be conceded, however, that there are cases in which the active euthanasia vs. passive euthanasia contrast genuinely does differentially affect our *moral judgments about a moral agent's character*. For example, it is clearly the case that doctors who quietly practice passive euthanasia at the rational request of their patients or their families are instances of one kind of moral personality, and that Dr. Jack Kevorkian—aka "Dr. Death"—was an instance of an altogether different kind of moral personality, namely, that of a "true believer" or moral fanatic. Nevertheless, this genuine difference in moral

judgments about character is morally non-basic, since it does not itself determine moral permissibility or impermissibility. But on the other hand, obviously, in some cases this difference in moral judgments is extremely important personally or socially. As a result of his practices of active euthanasia, and his explicit, published views on this, Dr. Kevorkian spent eight years in jail and was a highly controversial public figure, loudly criticized and even hated by many. By contrast, most doctors who quietly practice passive euthanasia at the rational request of their patients or their families are solid, successful, trusted citizens living ordinary, quiet lives.

A genuinely basic distinction in this area, however, is between

> (i) *voluntary euthanasia*, which is widely held to be sometimes morally permissible and perhaps sometimes also morally obligatory,
> (ii) *non-voluntary euthanasia*, which, similarly, is widely held to be sometimes morally permissible and perhaps sometimes also morally obligatory, and
> (iii) *involuntary euthanasia*, which is widely held to be morally impermissible.

What, more precisely, is this distinction? The answer has three parts.

First, by "voluntary euthanasia," I mean this:

> A mercy-killing *X* is voluntary euthanasia if and only if *X* follows from the actual or possible rational request of the real person who is mercy-killed, to be mercy-killed.

Second, by "non-voluntary euthanasia," I mean this:

> A mercy-killing *X* is non-voluntary euthanasia if and only if *X* follows from the merely possible rational request of the real person who is mercy-killed, to be mercy-killed, in cases in which that real person is temporarily or permanently unable to make an actual rational request, either because her rational capacities are temporarily or permanently offline (for example, in sleep, in temporary unconsciousness, in a coma, in a seizure, etc.), or because she is under some sort of informational blackout, misinformed condition, preventative constraint, or overwhelming internal or external compulsion or coercion (for example, she is unable to get accurate relevant medical information, or she is being given false medical information, or she is paralyzed with fear, or she is being threatened by a bad person, or a bad person is threatening to do something bad to other people if she does not accede to being mercy-killed, etc.).

Third and finally, by "involuntary euthanasia," I mean this:

> A mercy-killing *X* is involuntary euthanasia if and only if *X* occurs even despite an actual rational request by the mercy-killed real person *not* to be mercy-killed, that is, even despite an actual rational *refusal* by the mercy-killed real person to be mercy-killed (for example, via a "living will").

What, now, about the moral permissibility, impermissibility, or obligatoriness of voluntary euthanasia, non-voluntary euthanasia, and involuntary euthanasia? Before I can answer that question, we will need one preliminary definition. By *person-destroying suffering*, I mean suffering that is so intense, so prolonged, and so unrelievable that only death will prevent the higher-level or Kantian real human person permanently losing her rational or real-personal capacities altogether as a result of this suffering. Then, according to Existential Kantian Ethics and The Rage-Against-the-Dying-of-the-Light Theory of the nature and moral value of death, five specific first-order substantive ceteris paribus objective moral principles directly follow.

First, voluntary euthanasia is morally permissible if and only if

(ia) the real person who is actually or possibly rationally requesting to be mercy-killed is suffering, and
(ib) there is good reason to believe, although perhaps not overwhelmingly good reason to believe, that if this real person continued to live, then her suffering would then be personhood-destroying; otherwise voluntary euthanasia is morally impermissible.

Second, voluntary euthanasia is morally obligatory if and only if

(iia) voluntary euthanasia is morally permissible, and
(iib) there is overwhelmingly good reason to believe that if this real person continued to live, then her suffering would then be personhood-destroying; otherwise voluntary euthanasia is morally impermissible.

Third, non-voluntary euthanasia is morally permissible, morally obligatory, and morally impermissible under exactly the same set of conditions as voluntary euthanasia, mutatis mutandis, hence if and only if

(iiia) voluntary euthanasia is morally permissible, and
(iiib) there is overwhelmingly good reason to believe that if this real person continued to live, then her suffering would then be personhood-destroying; otherwise non-voluntary is morally impermissible.

Fourth, involuntary euthanasia is morally permissible if and only if

(iva) the real human person who is actually refusing to be mercy-killed is also actually suffering, and
(ivb) there is overwhelmingly good reason to believe that if this real person continued to live, then her suffering would then be personhood-destroying; otherwise involuntary euthanasia is morally impermissible.

Fifth and finally, involuntary euthanasia is never morally obligatory.

I will call the conjunction of these five specific first-order substantive ceteris paribus objective moral principles, collectively, *The Existential Kantian Ethics-Based Theory of Euthanasia*. The basic rationale behind The Existential Kantian Ethics-Based Theory of Euthanasia obviously follows from the now-familiar general first-order ceteris paribus objective moral principle that postulates

> (i) the impermissibility, other things being equal, of harming real persons by violating their dignity as real persons,

together with two equally familiar general first-order substantive ceteris paribus objective moral principles to the effect that,

> (ii) other things being equal, we ought to prevent or reduce dignity-violating harm to real persons, and
> (iii) we absolutely always ought to prevent or reduce the degradation of real persons.

These three moral principles combine to make it morally permissible to heed the actual or possible rational requests of any real human persons who are experiencing, quite likely will experience, or almost certainly will experience, personhood-destroying suffering by virtue of their continued life.

Moreover, and perhaps surprisingly, sometimes this inherently merciful intention to kill someone can even morally override a real human person's actual rational request not to be mercy-killed—which is to say that involuntary euthanasia is sometimes morally permissible. Looking around for possible examples, it may seem initially obvious that at least sometimes, overriding someone's actual rational refusal of euthanasia as expressed in what is legally known as a *living will* and then mercy-killing him even despite his express earlier intentions, is morally permissible. But living will cases are made complicated by the fact that there is necessarily a time-lag of some sort between the commission of the living will and the mercy-killing situation. This means that for many or perhaps even most such cases, it remains an open question whether, in the light of the changing cognitive circumstances and new information that always emerge over time, the mercy-killed real human person would have retroactively rationally revoked the earlier refusal of mercy-killing, were she to have retained roughly the same level of rationality as she did at the time when she made the statements made in the original will, and were also to have taken the changing circumstances and new information into careful consideration.[329]

Given this complexity, a special case that is at all not science-fictional and very likely to have actually happened in the history of war, perhaps many times, adapted from another famous paper by Foot,[330] is more rationally compelling. Suppose that a soldier deeply imbued with a sense of military honor is severely wounded and cannot be either sedated or moved by his rapidly retreating army in the face of a rapidly advancing army of enemies known to be cruel torturers, yet he adamantly refuses his comrades' offer to mercy-kill

him. Here it seems at least morally permissible for his comrades to override his rational request and mercy-kill him, for two reasons.

First, the wounded soldier's adamant refusal is morally equivalent to his rationally choosing either self-enslavement that ends in death, or suicide, in normal or everyday non-military contexts, both of which would be morally impermissible, other things being equal, precisely because these choices, although rational, are also self-harming acts that violate his own dignity as a rational human minded animal or real human person.

And second, almost certainly, the wounded soldier is going to suffer horribly at the hands of these moral monsters, to the point of personhood-destruction. Other things being equal, we are obligated to prevent or reduce dignity-violating harms to real persons, and also to prevent or reduce their degradation. In this context, killing the soldier is the only way that these moral principles can be followed. Hence in this context, due to ineluctably contingent factors of proximity, distance, temporality, and causation, as well as egocentrically-centered emotional relations and social relations, mercy-killing is clearly the lesser evil of the available options, and most keeps rational faith with the Categorical Imperative.

Therefore mercy-killing is morally permissible in this special case.

Something that is especially noteworthy about the morality of euthanasia is the role played by actual or possible *rational requests*. As I am understanding this notion, a rational request for something X by a rational animal or real person P is P's asking, with some legitimate warrant, to have X done for P's sake, and possibly also for the sake of others. It seems clear that a rational request by a real person P to have something X done entails a rational consent given by P to having X done for P's sake. But obviously P could rationally consent to X without having rationally requested it, or indeed without even wanting to request it rationally. So rational requests have more moral content than rational "consents"—with apologies for the neologistic pluralization—in the sense that rational requests more fully express a higher-level or Kantian real human person's capacity for principled authenticity.

This applies directly to the special case of the wounded soldier. Given the extra moral content of rational requests over and above rational consents, it seems clear that, other things being equal, we should always heed or at least seriously consider the actual or possible rational requests of others, and in particular we should always heed or at least seriously consider their actual rational requests not to be mercy-killed. To do otherwise would be *flagrant paternalism*, which undermines the paternalized higher-level or Kantian real human person's autonomy and thereby violates her dignity. *Non-flagrant paternalism*, by contrast, is *morally permissible paternalism*—for example, other things being equal, via the good parent-child relationship, the good teacher-student relationship, the good advisor-advisee relationship, or the good counselor-counselled relationship, etc. In such cases, there is an actual or possible rational request on the part of the non-flagrantly paternalized higher-level or Kantian real human person to be well-guided by the other higher-level or

Kantian real human person who plays the role of the good parent, teacher, adviser, counsellor, etc. In the special case of the wounded soldier case, in order to avoid flagrant paternalism, although it is morally *permissible* to mercy-kill the wounded soldier in this special set of circumstances, it cannot also be morally *obligatory* to mercy-kill him.

I must now address some possible critical worries about The Existential Kantian Ethics-Based Theory of Euthanasia.

The first worry is epistemological. Obviously in some cases, or perhaps even in a large number of cases, it will be very difficult to tell or to predict with perfect or even reasonable confidence whether someone's suffering really is, or really will be, personhood-destroying or not. These facts can be smoothly accommodated by The Existential Kantian Ethics-Based Theory of Euthanasia, however, by pointing out again (see chapter 2 above) that it is strictly the Existential Kantian Ethics-based, No-Foolish-Consistency-approach-driven, nonideal Kantian hierarchical structuralist theory of moral principles that is at issue here, and *not* the epistemology of moral judgment.

But these facts do also indirectly raise a deeper second critical worry, which is this: Why it is that only personhood-destroying suffering, and not other kinds of morally significant suffering—that is, other highly intense experiences of bodily or emotional pain—can justify the moral permissibility of euthanasia? Perhaps not too surprisingly, the answer has to do with what Existential Kantian Ethics takes to be the Highest or Supreme good, namely the achievement or realization of principled authenticity, at least partially or to some degree.

More precisely, the answer consists in making two distinct but closely related points.

First, the pursuit of principled authenticity by higher-level or Kantian real human persons necessarily requires their being alive. Hence it is only when *rational animality or real personhood itself* would be destroyed that euthanasia is morally permissible, because only that destruction will rule out altogether the minded animal's power for wholeheartedly pursuing principled authenticity. The highly intense experience of bodily pain is not alone sufficient, and will not, in and of itself, get you off the hook of trying to achieve principled authenticity. It remains true that sometimes the highly intense experience of bodily pain also constitutes personhood-destroying suffering. So obviously, it will depend heavily on the relevant higher-level or Kantian real human person herself, and also on the relevant actual context, whether a given highly intense experience of bodily pain yields personhood-destroying suffering, or not.

Second, since the suffering of higher-level or Kantian real human persons is always based on practical reasons, and since the suffering of such persons is always to some extent self-consciously or self-reflectively chosen, it always remains at least in principle possible for these persons, at any time, to achieve or realize principled authenticity, at least partially or to some degree. For you can choose, in a Rilkean, Sisyphean, Tractarian, or Nietzschean fashion, but also in a Kantian key, to enter the world of the "happy" (that is, in this context, the wholeheartedly autonomous) person. You *freely can* change your life, because you

morally must change your life, through the motive of respect, and for the sake of the Categorical Imperative. That is the existentially-pregnant formulation of the Kantian "*ought* entails *can*" principle. Only the destruction of the real human person herself decisively rules this out. So no condition of suffering short of personhood-destruction can morally justify euthanasia. Anything short of that would fail to respect the dignity of the person herself.

Nevertheless it remains true that, other things being equal, rational requests should be heeded, or at least seriously considered, in order to avoid flagrant paternalism. And as the special case of the wounded soldier clearly shows, sometimes these first-order substantive ceteris paribus objective moral principles will be in conflict. But here, as always in cases of conflicts of first-order substantive ceteris paribus objective moral principles, according to the No-Foolish-Consistency-approach-driven, nonideal Kantian hierarchical structuralist theory of moral principles,

the conflict of principles is automatically resolved in context and yields a single duty—assuming, of course that this is one of the act-contexts in which there *is* in fact a single duty applicable to it—via the practically constructive rational moral meta-procedures provided by The No-Global-Violation Constraint, The Excluded Middle Constraint, and The Lesser Evil Principle (see section 2.3 above). Nevertheless, also as always in cases of conflicting principles, the urgent agent-centered question of how to judge correctly and act rightly, in context and in the thick of things, then and there, it is not itself resolved by the logic of morality. For better or worse, then and there, the onus is on the agent herself.

6.8 THE MORALITY OF SELF-SACRIFICE

By *self-sacrifice* in the present connection, I mean what is commonly called the *supreme* or *ultimate* sacrifice, that is, intentionally sacrificing one's own rational minded animal or real personal life for the sake of something (say, a noble cause) or someone else (say, an innocent mortally threatened child). Merely dying in the course of saving someone else's life—for example, being accidentally hit by a car while carrying a vial of life-saving medicine across the street—is not self-sacrifice in this sense. In self-sacrifice, the inevitability or very high probability of one's own death in the very course of your promoting, protecting, and sustaining the highest or supereme moral value, or dignity, of something or someone else, is an inherent part of the intentional content of the act.

Even so, the supreme or ultimate sacrifice does not consist in biological or organismic death per se, however, since such a sacrifice can be made even if there is no organismic death. Thus if you were, for example, to agree to allow yourself to be reduced to a persistent vegetative state in order to save other people's lives by donating several vital organs to them, then that would still a supreme or ultimate sacrifice on your part. Or if you were to be reduced to such a state in the course of carrying out your risky duties as a fireman. Or

if you were to be reduced to such a state as a result of choosing to carry a normal, healthy third-trimester fetus to term when it was known that this would be extremely risky to you as the mother. And so-on. Then in each of these cases that would count as having made the supreme or ultimate sacrifice, despite the fact that an individual human animal bearing your proper name survived the act of self-sacrifice. So you did not sacrifice your human animal, but you did sacrifice your higher-level or Kantian real human personhood—your specifically *rational* human animality—and thus you sacrificed your very own life.

What makes it the *supreme* or *ultimate* sacrifice, moreover, is not the mere fact that you have given up your further opportunities for having subjective experiences. This can be shown by a set of conceivable and possible variants on the three cases (namely, donating vital organs, fireman, and birth mother) in which we hold the intentional motivations fixed but instead allow ourselves to be reduced to a conscious mental condition equivalent to permanent amnesia, or to a conscious mental condition equivalent to the final stages of Alzheimer's disease. Those, surely, would also count as cases of self-sacrifice. So it is not even the destruction of your capacity for *subjective experiences per se* that really matters, but instead the destruction of your capacity for *rational self-conscious consciousness, and free agency*. Otherwise and more explicitly put, what makes it the supreme or ultimate sacrifice is just the fact that you have intentionally given up, for the sake of something or someone else, *all of your further powers for and opportunities to pursue principled authenticity*.

The moral issue of self-sacrifice has already arisen in earlier chapters in at least five different contexts:

> (i) in the context of the boy's choice between his mother and joining the French Resistance against the Nazis, in Sartre's famous example (section 2.0),
> (ii) in the context of abortion and the conditions under which someone is permitted to refuse life-support to another real person, and in particular with respect to the unconscious cyclist example (section 4.3),
> (iii) in the context of the famous *Fat Man* case in The Trolley Problem (section 5.1),
> (iv) in the context of morally permissibly killing innocent attackers (section 5.2), and
> (v) in the context of the famous *Pond* case in The Singer-Unger Famine Relief Problem (section 5.3).

Correspondingly, in relation to all of these contexts, and according to Existential Kantian Ethics and The Rage-Against-the-Dying-of-the-Light Theory, three things seems clearly and distinctly true.

First, (A), there is a set of conditions under which self-sacrifice is morally permissible because it incorporates one or more first-order substantive ceteris paribus objective moral principles, and also because it would also be morally sublime to act in this way, although it is nevertheless supererogatory.

Second, (B), there is a set of conditions under which self-sacrifice is morally impermissible because it would thereby be a violation of someone's dignity as a rational minded animal or real person, whether someone else's dignity or one's own.

And third, (C), there is a set of conditions under which self-sacrifice is not only morally permissible but also morally obligatory, because either some special moral commitment or some special moral wrong makes it one's personal responsibility to lay down one's life, and there is no other way in this actual context that the relevant commitment can be met or the relevant wrong can be prevented or stopped, due to ineluctably contingent factors of distance, proximity, temporality, and causation, as well as egocentrically-centered emotional relations and social relations.

Let us now consider each of these cases in a little more detail.

Re (A): Self-sacrifice is clearly morally permissible in some cases, precisely because it flows from one or more of the first-order substantive ceteris paribus objective moral principles, other things being equal,

(i) not to harm others by violating their dignity,
(ii) to prevent or reduce dignity-violating harms to others,
(iii) to prevent or reduce the degradation of others, and
(iv) to promote the happiness of others.

For example, a mother can morally permissibly choose to provide life-support to a normal, healthy neo-person even when it means that she herself will die. So too the Fat Man can permissibly choose to throw himself down from the bridge onto the tracks in front of the runaway trolley in order to save five other people in *Fat Man*. And I can morally permissibly choose to sacrifice myself by becoming the Fat Man's life-saving landing pad in either *Well-Armed Innocent Attacker* or *Well-Armed Defensive Attacker*. Each of these choices and acts is morally sublime, if chosen or done for the sake of the Categorical Imperative and for the sake of other higher-level or Kantian real human persons in The Realm of Ends, who are then all regarded, considered, and in this case also all equally treated as absolutely intrinsically valuable ends-in-themselves.

But at the same time, each of these choices or acts is also significantly more than is morally required in that context. *Equal consideration* of others in The Realm of Ends does not automatically entail *equal treatment* of others in The Realm of Ends. Thus it would also be morally permissible to refrain from choosing or doing it, in view of the self-regarding first-order substantive ceteris paribus objective moral principles

(i) to promote one's own deep happiness and
(ii) to engage in self-perfecting projects,

other things being equal. It remains really possible that deep happiness or principled authenticity is partially or fully achievable or realizable in a way that does not require your self-sacrifice. Only if your deep happiness or principled authenticity constitutively depends on your self-sacrifice, would it be morally obligatory to lay down your own life for the sake of something or someone else. I will come back to this crucial point again shortly.

Re (B): The set of conditions under which self-sacrifice is impermissible must adequately reflect the content of the following set of five first-order substantive ceteris paribus objective moral principles specifically pertaining to self-sacrifice:

> (i) Foolhardy self-sacrifice is morally impermissible, other things being equal. For example, it is morally impermissible to lay down one's life for the sake of something or someone else, merely in order to enjoy the adrenaline rush of facing death.
> (ii) Glory-seeking self-sacrifice is morally impermissible, other things being equal. For example, it is morally impermissible to lay down one's life for the sake of someone or something else, merely in order to bring about the posthumous reputation of being a hero or martyr.
> (iii) Notoriety-seeking self-sacrifice is morally impermissible, other things being equal. For example, it is morally impermissible to lay down one's life for the sake of someone or something else, merely in order to bring about the posthumous Warholesque ten minutes or longer of fame that results from it.
> (iv) Manipulative self-sacrifice is morally impermissible, other things being equal. For example, it is morally impermissible to lay down one's life for the sake of something or someone else, merely in order to coerce or force someone into choosing or doing or feeling something.
> (v) Hatred-driven or revenge-driven self-sacrifice is morally impermissible, other things being equal. For example, it is morally impermissible to lay down one's life for the sake of something or someone else, merely in order to cause them either to experience bodily pain or to suffer.

It is clearly and distinctly true of each of these principles, and also of its corresponding example, that some higher-level or Kantian real human person's dignity is being violated by the self-harming act of self-sacrifice, whether one's own dignity (as in (i)-type cases, (ii)-type cases, and (iii)-type cases) or someone else's dignity (as in (iv)-type cases and (v)-type cases). This feature, in turn, provides a necessary and sufficient condition for morally impermissible self-sacrifice:

> A self-sacrifice *X* is morally impermissible, other things being equal, if and only if *X* thereby harms some rational human animal or real human person by violating her dignity.

Re (C): Self-sacrifice is morally obligatory if it flows directly, as a matter of personal responsibility, from some special and morally sublime commitment—such as a moral commitment to the well-being of one's loved ones, family, and close friends, for the sake

of love, over and above the call of mere duty; or a moral commitment to a given morally sublime practice or a way of life; or a moral commitment to an extremely high-minded conception of honor; or a moral commitment to protect others above and beyond the call of mere duty; or a moral commitment to rescue others above and beyond the call of mere duty, and so-on. —Provided that there is no other way in this actual context that the relevant moral commitment can be met, due to ineluctably contingent factors of distance, proximity, temporality, and causation, as well as egocentrically-centered emotional relations and social relations. For example, it is morally obligatory to lay down your life for the sake of your beloved husband, wife, or partner, your children or siblings, or your closest friend, if, in context, awful push comes to awful shove. And self-sacrifice also is morally obligatory if it flows directly, as a matter of personal deep moral responsibility, from the fact that it is the only way of your preventing or reducing the impact of some irremediably or irreparably terrible moral wrong that you yourself have brought about or directly done, and again there is no other way in this context that the relevant commitment wrong can be prevented or stopped, due to ineluctably contingent factors of distance, proximity, temporality, and causation, as well as egocentrically-centered emotional relations and social relations. In both kinds of case, your principled authenticity constitutively depends on your self-sacrifice. Hence it is morally obligatory. More generally then, for any real human person like us, self-sacrifice is morally obligatory if and only if that person's principled authenticity constitutively depends on her self-sacrifice.

Otherwise put, in cases in which self-sacrifice is morally obligatory, what is on the line, what is poised on the edge of a Kierkegaardian abyss of one thousand fathoms, is your achievement or realization of principled authenticity partially or to some degree. Otherwise put, it is your *integrity* that is on the line. Self-sacrifice in all such cases is therefore an absolutely intrinsically good thing—it has, in Kant's terms, moral worth—for the higher-level or Kantian real human person who lays down her life. Hence death_p is sometimes not merely an intrinsically good thing, but sometimes also even the highest or supremely intrinsically good thing, for the one who dies.

6.9 THE MORALITY OF SUICIDE

In section 5.3 above, I briefly argued that, other things being equal, I should not commit suicide except to prevent or reduce my own personhood-destroying suffering. Hence merely hating my own life, or merely suffering intensely, is not sufficient to justify suicide. This is precisely because at any time, no matter how awful and how miserable my life has been, as long as I am still a higher-level or Kantian real human person, I can always freely choose "to change my life," and achieve or realize principled authenticity at least partially or to some degree. I now want to unpack this line of reasoning further.

Self-sacrifice, as I have said, is intentionally laying down one's own real personal life for the sake of something or someone else. Suicide, by contrast, is intentionally killing oneself for one's own sake. Therefore, self-sacrifice and suicide are obviously conceptually and logically distinct from one another, and both of them in turn are conceptually and logically distinct from self-killing per se. For example, someone could accidentally kill himself, or be coerced, forced, or tricked into killing himself, and these clearly would not count as either self-sacrifice or suicide.

Nevertheless, there are at least two important parallels between the morality of self-sacrifice and the morality of suicide.

The first important parallel is that for each of the basic cases under which self-sacrifice is clearly morally impermissible, there is a direct corresponding analogue case for suicides that is also clearly morally impermissible, as follows:

> (i*) Foolhardy suicide is morally impermissible, other things being equal. For example, it is morally impermissible to kill oneself for one's own sake, merely in order to enjoy the adrenaline rush of facing death.
> (ii*) Glory-seeking suicide is morally impermissible, other things being equal. For example, it is morally impermissible to kill oneself for one's own sake, merely in order to bring about the posthumous reputation of being a hero or martyr.
> (iii*) Notoriety-seeking suicide is morally impermissible, other things being equal. For example, it is morally impermissible to kill oneself for one's own sake, merely in order to bring about the posthumous Warholesque ten minutes or longer of fame that results from it.
> (iv*) Manipulative suicide is morally impermissible, other things being equal. For example, it is morally impermissible to kill oneself for one's own sake, merely in order to coerce or force someone else into choosing, doing, or feeling something.
> (v*) Hatred-driven or revenge-driven suicide is morally impermissible, other things being equal. For example, it is morally impermissible to kill oneself for one's own sake, merely in order to cause someone else either to experience bodily pain or to suffer.

In precise analogy to the morality of self-sacrifice, then, it is clearly true of each of these moral principles and its corresponding example that some higher-level or Kantian real human person's dignity is being violated by the self-harming act of suicide,[331] whether one's own dignity (as in (i*)-type cases, (ii*)-type cases, and (iii*)-type cases) or someone else's dignity (as in (iv*)-type cases and (v*)-type cases). Hence again in precise analogy to the morality of self-sacrifice, this feature in turn provides a necessary and sufficient condition for morally impermissible suicide:

> A suicide *X* is morally impermissible, other things being equal, if and only if *X* thereby harms some higher-level or Kantian real human person by violating her dignity.

The second important parallel between the morality of self-sacrifice and the morality of suicide is that each can be supported by the first-order substantive ceteris paribus objective moral principle which says that, other things being equal, we must prevent or reduce dignity-violating harms to all real human persons, specifically including all higher-level or Kantian real human persons. In the case of self-sacrifice it is dignity-violating harm *to others* that is morally salient, whereas in the case of suicide, it is dignity-violating harm *to oneself* that is morally salient. This in turn generates a reflexive modal criterion for morally permissible suicide, which can be rationally reconstructed by answering the following question:

> What kind of first-person harm or suffering is such that, other things being equal, merely failing to kill oneseself in order to alleviate that harm or suffering would always and necessarily thereby harm oneself by violating one's own dignity?

It is clear that only harm or suffering that is *personhood-destroying* meets this very high reflexive modal standard. Therefore, suicide is morally permissible if and only if continued life for the higher-level or Kantian real human person who commits suicide would involve harm or suffering that is personhood-destroying. Hence, other things being equal, we should always try to dissuade or stop people from committing suicide unless it is very clear that their continued life would be personhood-destroying, just as, other things being equal, we should always try to dissuade or stop people from enslaving themselves, or imprisoning themselves, because these reflexively-harming acts violate their own dignity as real persons.

The central role of the concept of personhood-destroying suffering in the morality of death can in some cases lead to an unexpected convergence of morally permissible euthanasia, morally permissible self-sacrifice, and morally permissible suicide. For example, in the year 2000, the 86 year-old Kantian philosopher Stephan Körner and his 79 year-old wife Edith, who was a National Health Service expert in the UK, committed double-suicide by taking a lethal overdose, then tying plastic bags around their heads and putting pillows on top of them.[332] She was suffering from terminal cancer, and they died together in each other's arms. Let us reasonably suppose that

> (i) Edith's suffering was person-destroying and she knew this,
> (ii) Stephan knew this too,
> (iii) they had both carefully thought through the moral implications of their double-suicide, and had jointly rationally consented to it,

and let us also suppose that

> (iv) Stephan first tied the bag around his wife's head and then around his own,

> (v) Stephan's emotional commitment to his wife was of such depth and intensity that he could not bring himself to kill her, or to assist her killing herself, unless he killed himself too,
> (vi) Stephan also knew that at his age he would not have been able to live on without his wife without also suffering in a personhood-destroying way,

and finally let us also suppose that

> (vii) they both knew very well the possible adverse legal implications of assisted suicide, hence they did not want to involve anyone else in their collective act.

Then by helping to give Edith a lethal overdose and then smothering her, Stephan morally permissibly mercy-killed her, and by his also taking a lethal overdose and smothering himself he morally permissibly sacrificed himself for her sake, and by his doing these two things they both committed morally permissible double-suicide.

Granting these suppositions, it also seems to me really possible that Stephan and Edith both achieved principled authenticity, at least partially or to some degree, by means of this collective act, and that this was a real-world case in which death was a supremely good thing for the real persons who died, at the very least morally permissible in that special context, and perhaps even morally obligatory in that special context. If so, then in that special context, by committing double suicide, they died with dignity. In any case, I will frankly admit to being deeply moved by what they did. Strikingly and significantly, however, their surviving daughter implicitly sharply disagreed with my moral analysis. She thought that it was clearly *morally impermissible* for her father Stephan to have committed suicide *too*:

> [Their daughter,] Dr Ann Altman, says she was "disappointed" at her father's actions and would be horrified if her parents' final act became lauded as the ultimate symbol of devotion. "I would want to leave my own children a different legacy," she says simply.[333]

And I can certainly morally empathize with her point of view—what if Stephan and Edith had been *my own parents*: how would I feel? Moreover, what would be the precise rational bearing of that feeling on my moral (im)partiality? Hard, subtle questions! I will "leave it as a task for the reader" to reflect further on the moral complexities of this poignant case.

6.10 THE MORALITY OF ONE'S OWN ACCIDENTAL DEATH$_P$

By *an accidental death$_p$* I mean a death$_p$ that is

> either (i) for all practical intents and purposes, *rationally unforeseeable*, in the sense that its knowable actuarial probability is extremely low [type-(i)],
> or (ii) whose knowable actuarial probability, while relatively high, is such that even though it is not entirely rationally unforeseeable, the type of situation in which it actually occurs is no more likely to be reasonably regarded as unusually risky than many other types of situations reasonably regarded as entirely ordinary and relatively unrisky [type-(ii)],
> or (iii) whose knowable actuarial probability is very high, and such that intentionally engaging in practices that involved such situations would be generally regarded as highly risky behavior, although in fact many or even most people who engage in those practices are not in fact killed by doing so [type-(iii)].

In other words, in my sense of "accidental death$_p$," the following would *all* count as accidental deaths$_p$:

> (a) dying by being struck by lightning, by being in an airplane accident, or by falling off the back of a moving train—as in *Double Indemnity*[334] (type-(i) cases),
> (b) dying in a car accident while driving at night on a US interstate highway, or dying in middle age from some form of cancer (type-(ii) cases), and
> (c) dying during an attempt to fly around the world by oneself, like Amelia Earheart, or dying during an ascent of Mount Annapurna or K2 (type-(iii) cases).

By contrast, being killed-in-action during an armed crime, or during a war, would not count as accidental deaths but instead count as what I will call *not-unexpected deaths*. The basic idea behind not-unexpected deaths is that not only is it the case that the knowable actuarial probability of dying while engaging in practices that involve such situations is very high, but also one could reasonably expect to die as a result of engaging in them.

Many or perhaps even most cases of accidental death are commonly said to be "tragic." Now the specifically modern, as opposed to classic Greek and Aristotelian, strict literary, and moral sense of "tragic," vividly exemplified in the case of Hamlet, as I mentioned earlier in passing, typically involves the actuality or real possibility of greatness of character in a certain real human person like us, a correspondingly great character flaw in that real person, a terrible downfall for that real person as a direct result of that great character flaw, and some sort of cathartic experience for the witnesses of this downfall. In fact, very few accidental deaths$_p$ really are such a thing, no matter how catastrophic and unfortunate they are for the real person who died, for her loved ones, her close friends, her co-workers, etc. But at the same time, it does seem to be true that all or almost all cases of accidental death$_p$ are such that they are an inherently bad thing for the higher-level or

Kantian rational human minded animal who dies, in the sense that, by ending a life that has not yet manifested the achievement or realization of principled authenticity, at least partially or to some degree, they inherently fall short of that High-Bar normative standard. They are therefore *untimely* deaths.

Two further important moral questions arise here.

First, can there ever be any accidental deaths$_p$ *that are also timely deaths$_p$*, involving something inherently good for the higher-level or Kantian real human person who accidentally dies?

Second, in view of the universal, or almost universal, inherent badness of accidental death$_p$—in that always, or almost always, by ending a life that has not yet manifested the achievement or realization of principled authenticity, at least partially or to some degree, it inherently falls short of that High-Bar normative standard—how ought we then to think about the much-greater-than-merely-non-zero probability *that we ourselves shall die an accidental death$_p$*?

As regards the first question, it is logically, metaphysically, and naturally possible that some accidental deaths$_p$ are such that they also accidentally prevent or reduce personhood-destroying suffering, or accidentally prevent or reduce the impact of some irremediably heinous act. For example, it is logically, metaphysically, and naturally possible that just as someone is about to begin a protracted process of personhood-destroying suffering, or is just about to commit some irreparably terrible sin, she is killed in a car accident. Such an accidental death would be inherently good for the higher-level or Kantian real human person. But presumably that is very rare indeed, and in the nature of things it would be simply a matter of good moral luck.

Moreover, it is very hard to see how an accidental death$_p$ could ever also be a *supremely* good thing for that person. For example, it is possible that just as someone achieves principled authenticity partially or to some degree, she is killed in a car accident, or drowns. There appears to be no way in which such a death$_p$ can be *inherent* to her principled authenticity. Such a death$_p$ neither *undermines* her principled authenticity nor in any way *constitutes* her principled authenticity. Although this sort of accidental death$_p$ does indeed satisfy Death's Excluded Middle in that, by hypothesis, her death$_p$ positively manifests principled authenticity, nevertheless it is morally *otiose* or *pleonastic*. As such, this kind of accidental death$_p$ is neither good nor bad *for* the person who dies, although it may of course have significant positive or negative moral value, as a good or bad fact *about* that person, in many extrinsic relational or third-person ways—for example, in its impact on those lives have been directly or indirectly benefitted or bettered by her actions, on those who love her, her close friends, her co-workers, etc.

As regards the second question now, it seems to me that the much-greater-than-merely-non-zero probability of accidental death$_p$, if regarded as providing a sufficient reason *not* to pursue principled authenticity, would itself be morally self-stultifying and indeed morally impermissible. This is simply because, as long as higher-level or Kantian real

human persons *are* alive-and-kicking,[335] then they do have sufficient reason to pursue principled authenticity, since achieving it at least partially or to some degree, is the Highest or Supreme Good. From this it follows that, other things being equal, one ought to choose and act as if the possibility of accidental death were negligible, whenever one is acting according to some or another first-order substantive ceteris paribus objective moral principle, in pursuit of principled authenticity. Here I have specifically in mind, for example, the principle which says that, other things being equal, we ought to promote our own happiness and engage in self-perfecting projects; or the principle which says that, other things being equal, we ought to prevent or reduce dignity-violating harms; or the principle which says that, other things being equal, we ought to prevent or reduce degradation. I will call this more general principle *The Reasonable Bravery Principle*.

In other words, according to The Reasonable Bravery Principle, a non-trivial but still not excessive level of courage, other things being equal, is morally obligatory as an inherent concomitant of following other first-order substantive ceteris paribus objective moral principles in pursuit of principled authenticity. Contrapositively, other things being equal, "excessively risk-averse" choice and action, namely, cowardice, is morally impermissible whenever one is obligated by other first-order substantive ceteris paribus principles, in pursuit of principled authenticity. It seems obvious, but is probably worth explicitly noting, that merely feeling fear, even intensely feeling fear, is not the same as cowardice. Cowardice, as I mentioned just above, is excessively risk-averse choice or action. But as has been many times pointed out, overcoming fear, perhaps even intense fear, is an essential part of courage, where courage is, as Aristotle very correctly, if somewhat tautologously, points out in *The Nicomachean Ethics*, the virtue which consists in being brave to an appropriate extent in all the relevant situations that manifestly require bravery.

Correspondingly, however, it not tautologous to note that The Reasonable Bravery Principle adequately captures some of the basic moral content of the Aristotelian virtue of courage. All people who wholeheartedly follow The Reasonable Bravery Principle will actually be courageous in Aristotle's sense. Obviously, there can be other sorts of courage as well: for example, courage in the face of possible severe criticism by others or in the face of public embarrassment, courage in the face of the possible failure of one's own deep-happiness-achieving or self-perfecting projects, and so on. But the crucial point here is that unlike the moral principles of Aristotelian virtue ethics, the Existential Kantian Ethics-based moral principles like The Reasonable Bravery Principle are always substantive and synthetic a priori, not tautologous and analytic.

6.11 THE MORALITY OF ONE'S OWN NATURAL DEATH$_P$

Finally, we arrive at the moral-existential sticking-point. By *a natural death$_p$*, I mean a death$_p$ that is neither the result of mercy-killing, nor the result of self-sacrifice, nor the result of suicide, nor an accidental death$_p$, nor a not-unexpected death$_p$. The prime example of a natural death$_p$ is dying in old age from the deleterious natural effects of aging. This can include dying from one or more of the same causes that, at an earlier stage in one's life, would have classified a death$_p$ as accidental—for example, diseases such as cancer, or a heart-attack. So the "naturalness" of a natural death$_p$ is determined, in part, relative to the normal life-expectancy for real human persons like us under the particular environmental, historical, and social conditions obtaining in that context.

Now given Death's Excluded Middle, all natural deaths are either timely and inherently good for the person who dies (aka "dying with dignity") or else untimely and inherently bad for the person who dies, and never both. Moreover, a natural death ND_p is timely if and only if

> either (i) ND_p is an inherently good thing for the higher-level or Kantian real human person who dies, because continued life would be personhood-destroying for her,
> or (ii) ND_p is not only an inherently good thing for the higher-level or Kantian rational human animal who dies, but also a *supremely* good thing for her, because by means of the process of dying she thereby achieves or realizes principled authenticity, at least partially or to some degree.

Otherwise, all other natural deaths are untimely. Or in other words, all natural deaths for lives in which principled authenticity has *not* been manifested in *any* way are inherently bad for the real person who dies, precisely because she has thereby failed to satisfy the high-bar moral norm of achieving or realizing principled authenticity, at least partially or to some degree. In short, in a moral sense, such lives have been *wasted*.

Granting that, and taking a realistic but not cynical view of rational "human, all-too-human" nature, and of the rational human condition, it seems very likely that in the natural course of things, sadly, *a great many natural deaths$_p$ have been, are, and will be untimely*. In this way, a very obvious but also very important moral question arises at the egocentric center of The Web of Mortality: Should I—should we—fear an untimely natural death$_p$?

My answer to this question, from the standpoint of Existential Kantian Ethics and The Rage-Against-the-Dying-of-the-Light Theory of the nature and moral value of death, is: *No*, we ought not to fear an untimely natural death. There are three reasons for this.

First, by The Reasonable Bravery Principle, reasonable bravery is morally obligatory with respect to all accidental deaths$_p$. But natural death$_p$ adds nothing to accidental death$_p$

that would give us a new sufficient moral reason for fear. Hence reasonable bravery is also morally obligatory with respect to all natural $deaths_p$, including the untimely ones.

Second, although it is true, as both Aristotle and also Nagel have correctly argued,[336] that it is possible for people to be harmed in a extrinsic relational sense after their natural $deaths_p$ and during the finite or sempiternal time of their $deaths_s$—for example, by the post-mortem revelation of awful secrets about them, by the bad post-mortem consequences of their choices or acts, or by the post-mortem misfortunes of their loved ones, friends, families, or larger social communities, etc.—nevertheless this is always something that is only ever a bad thing *about* them, from the third-person point of view, and never something that is a bad thing *for* them, from the first-person point of view. Intrinsic or first-personal harms require a living higher-level or Kantian real human person who is harmed in the actual course of her real human personal life-process, that is, in the actual spatiotemporal and causal sequence of her complete, finite, and unique life.

Furthermore, the only intrinsic, first-personal harms that we morally have sufficient reason to fear are those that harm us by violating our dignity. Since, like all real persons, all higher-level or Kantian real human persons are literally identical with their complete, finite, and unique life-processes or lives, then they have dignity just as long as they are alive, and at no other times. Hence higher-level or Kantian real human persons cannot be harmed by violating their dignity *after* their natural $deaths_p$, hence during the finite or sempiternally infinite time of their $deaths_s$. And for the same reason, they cannot be harmed by violating their dignity *before* they are born. This moral fact about us is quite easy to see with respect to the natural time prior to the beginning of our lives, when we did not yet exist; but the very same moral fact applies just as much to the finite or sempiternally infinite time following our own deaths, when we no longer exist.

So on the one hand, in this specific regard Lucretius was absolutely right: There is indeed at least one metaphysical and moral symmetry or mirroring between the time prior to our births and the finite or sempiternally infinite time during our $deaths_s$,[337] in that we cannot be intrinsically morally harmed during either time. Therefore we should not fear being intrinsically morally harmed *after* our own untimely natural $deaths_p$, any more than we do or should fear being intrinsically morally harmed *before* our lives begin. On the other hand, however, as I noted earlier, Lucretius was as it were "dead wrong" about the symmetry or mirroring thesis with respect to $death_p$. The pre-natal non-existence of a higher-level or Kantian real human person is essentially different from her $death_s$, precisely because her $death_s$ is necessarily post-life, and therefore it inherently presupposes her actual $death_p$, whereas his pre-natal non-existence is necessarily not post-life, and therefore it does not metaphysically include her actual $death_p$.

Third, and most importantly of all, the second individually sufficient (but not individually necessary) condition for a timely natural $death_p$ *ND* says that *ND* is not only an inherently good thing for the higher-level or Kantian rational human minded animal who dies, but also a *supremely* good thing for her, because by means of the process of

dying she achieves principled authenticity, at least partially or to some degree. What I want specifically to highlight with respect to this second criterion is that it is really possible to achieve principled authenticity, at least partially or to some degree, *even only at the very end of one's life*, by means of dying a natural death_p. One way of seeing this is to double-underline a remarkable *moral-existential-bootstrapping* feature of the pursuit of principled authenticity. If you really and truly are wholeheartedly trying to achieve or realize principled authenticity, at least partially or to some degree, then you are thereby *already* really and truly achieving or realizing principled authenticity, at least partially or to some degree. And your natural death_p cannot change this essential moral fact about you and your life. Indeed, for someone who is really and truly wholeheartedly trying to achieve or realize principled authenticity, at least partially or to some degree, his natural death_p is precisely the intrinsic closure of such an inherently morally good life, at least to that extent. Therefore it is a timely natural death_p, and "dying with dignity."

Here I am not talking about "Stoicism," as that notion is commonly understood. It seems to me self-evidently true that if one were to achieve principled authenticity even , at least partially or to some degree, even only at the very end of one's life, by dying a natural death_p, then a proper part of this achievement would *not* be to accept the beginning of one's own death_s with passive and emotionless rational resignation in the face of overwhelming natural forces, but on the contrary to affirm both one's own natural death_p and also one's own death_s wholeheartedly as the intrinsic closure of one's own complete, finite, and unique life, and the terminating form or immanent structure of one's own life. What is needed, then, is a thoroughly *active and passionate Kantian Stoicism*. Furthermore, as should be obvious by now, it also seems to me that the moral-emotional core of this thoroughly active and passionate Kantian Stoicism about death_p and death_s alike, is captured precisely by Dylan Thomas's famous poetic rant, at once Dionysian and Thanatosian:

> Do not go gentle into that good night,
> Old age should burn and rave at close of day;
> Rage, rage against the dying of the light.
>
> Though wise men at their end know dark is right,
> Because their words had forked no lightning they
> Do not go gentle into that good night.
>
> Good men, the last wave by, crying how bright
> Their frail deeds might have danced in a green bay,
> Rage, rage against the dying of the light.

Wild men who caught and sang the sun in flight,
And learn, too late, they grieved it on its way,
Do not go gentle into that good night.

Grave men, near death, who see with blinding sight
Blind eyes could blaze like meteors and be gay,
Rage, rage against the dying of the light.

And you, my father, there on the sad height,
Curse, bless me now with your fierce tears, I pray.
Do not go gentle into that good night.
Rage, rage against the dying of the light.[338]

I will assume, now, that The Rage-Against-the-Dying-of-the-Light Theory of the nature and moral value of death is true. It follows that since our own states of being dead, our deaths_s, inherently cannot be subjectively experienced, then at the very moment of death_p, our process of dying, which brings us up to the very beginning of our permanent deaths_s, we will still be alive and subjectively experiencing. Now also suppose also that at that time we are lucky enough to have suffered no personhood-destroying accident or disease, and are also still higher-level or Kantian real human persons, in possession of our basic capacities for intentionality, caring, and rationality. In all such cases, then even if someone has not yet achieved or realized principled authenticity *at all*, nevertheless there is always enough time left for her wholeheartedly to affirm her own natural death as the intrinsic closure of her own complete, finite, and unique life (or more generally, wholeheartedly to choose or do something or another for the sake of any of her own moral principles and the Categorical Imperative), since this can be chosen at any time right up to and including the very moment of the beginning of her own death_s. In so choosing or so doing, she can thereby achieve principled authenticity, at least partially or to some degree, by freely conferring timeliness and "dying with dignity" on her own natural death_p.

In this way, seemingly paradoxically, even only at the very end of your life, *your own natural* $\textit{death}_p$ *can also be a way of changing your life*. And if we can achieve principled authenticity, at least partially or to some degree, at any time right up to and including the very moment of the beginning of our own deaths_s, by freely conferring timeliness and "dying with dignity" on them, by changing our lives, and by converting them from ongoing projects into completed projects, like finishing a book or creating a work of art, then there is no sufficient moral reason for us to fear our own untimely natural deaths_p. For every such natural death_p will necessarily be timely and dignified, not untimely and undignified.

On the contrary, then, there is instead a sufficient moral reason for each and every one of us wholeheartedly to affirm his own natural death_p as the intrinsic closure of his complete, finite, and unique life, or more generally, wholeheartedly to choose or do something or another for the sake of any of her own moral principles and the Categorical

Imperative, through respect for the dignity of real persons, whether others' dignity or one's own, at any time right up to and including the very moment of the beginning of his $death_S$, provided that it constitutes a genuine change-of-heart. So no matter how *wrong* everything else has been in your life, as long as you are still alive, sentient, and sapient, then there is always enough time left for getting it at least partially or to some degree *right*.

6.12 CONCLUSION

A certain kind of rational human life—a life in which principled authenticity is achieved or realized, at least partially or to some degree—is truly worth living. Indeed, if I am correct, then it is the *only* kind of rational human life that is truly worth living. And of course we do not live ideally, in a void, or alone, in this world of ours. So the meaning of a rational human life is *the pursuit of principled authenticity, in solidarity with all other real human persons and alongside all other minded animals, everywhere, in this thoroughly nonideal natural and social world*.

But in order to have such a life, we must live wholeheartedly for the sake of all and only those things that are truly worth dying for; and all of them are inherently bound up with respect for the nondenumerable absolute intrinsic objective value, or dignity, of real human persons. This may seem paradoxical, *living for just those things truly worth dying for*, but it is not.

It is built into the nature of the rational human condition, built into the nature of our complete, finite, and unique lives, and therefore also built into the morality of our own deaths. Therefore do not go gentle into that good night. On the contrary, you ought to rage against the dying of the light. And because you ought to do it, it follows that you freely can. So in this way, by thinking about the morality of our own *deaths*, we have come all the way around, yet again, to Rainer Maria Rilke's terse and intensely beautiful formulation of the Categorical Imperative, in terms of our own *lives*: *Du musst dein Leben ändern*.[339] You must change your life.

REFERENCES

[1] S. Kierkegaard, "Purity of Heart is to Will One Thing," in *The Essential Kierkegaard*, trans. H. Hong and E. Hong (Princeton, NJ: Princeton Univ. Press, 2000), p. 271.

[2] See also, for example, R. Louden, *Kant's Impure Ethics: From Rational Beings to Human Beings* (Oxford: Oxford Univ. Press, 2000); R. Louden, *Kant's Human Being: Essays on His Theory of Human Nature* (Oxford: Oxford Univ. Press, 2011); and P. Frierson, *What is the Human Being?* (London: Routledge, 2013).

[3] See Hanna, *Deep Freedom and Real Persons: A Study in Metaphysics*, esp. chs. 1-2.

[4] See G.E. Moore, *Principia Ethica* (Cambridge: Cambridge Univ. Press, 1903), esp. pp. 40, 58, and 73. See also section 1.4 below.

[5] See, for example, O. O'Neill, *Constructions of Reason* (Cambridge: Cambridge Univ. Press, 1989); T. Hill, *Dignity and Practical Reason in Kant's Moral Theory* (Ithaca, NY: Cornell Univ. Press 1992); B. Herman, *The Practice of Moral Judgment* (Cambridge, MA: Harvard Univ. Press, 1993); M. Baron, *Kantian Ethics (Almost) without Apology* (Ithaca, NY: Cornell Univ. Press, 1995); C. Korsgaard, *Creating the Kingdom of Ends* (Cambridge: Cambridge Univ. Press, 1996); C. Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge Univ. Press, 1996); R. Audi, *The Good in the Right* (Princeton, NJ: Princeton Univ. Press, 2004); A. Wood, *Kantian Ethics* (Cambridge: Cambridge Univ. Press, 2007); C. Korsgaard, *Self-Constitution: Agency, Identity, and Integrity* (Oxford: Oxford Univ. Press, 2009); D. Parfit, *On What Matters* (2 vols., Oxford: Oxford Univ. Press, 2011); and R. Audi, *Means, Ends, and Persons* (Oxford: Oxford Univ. Press, 2015).

[6] See, for example, B. Williams, *Ethics and the Limits of Philosophy* (London: Fontana, 1985); and B. Williams, *Morality: An Introduction to Ethics* (Cambridge: Cambridge Univ. Press, 1972). The ethics vs. morality = *Sittlichkeit* vs. *Moralität* contrast has also had some impact in contemporary philosophy. For example, essentially the same distinction is replicated in the titles and basic topics of the first two divisions of Russ Shafer-Landau's widely-used and influential *Fundamentals of Ethics* (Oxford: Oxford Univ. Press, 2015): "The Good Life" and "Normative Ethics: Doing the Right Thing," which sets it interestingly apart from the erstwhile bog-standard tripartite division of moral philosophy into *meta-ethics*, *normative ethics*, and *applied ethics*.

[7] See J.C. Calhoun, "Speech on the Reception of Abolition Petitions: Revised Report," *U.S. Senate* (Feb. 6, 1837, at Wake Forest University), available online at URL = <http://users.wfu.edu/zulick/340/calhoun2.html>.

[8] Williams, *Ethics and the Limits of Philosophy*, ch. 10.

[9] See, for example, F. Nietzsche, *Beyond Good and Evil*, trans. W. Kaufmann (New York: Vintage, 1966); F. Nietzsche, " The Genealogy of Morals," in F. Nietzsche, *The Genealogy of Morals and Ecce Homo*, trans. W. Kaufmann (New York: Vintage, 1967), pp. 13-163; M. Foucault, *Discipline and Punish: The Birth of the New Prison*, trans. A. Sheridan (New York: Vintage, 1975); M. Foucault, *The Order of Things* (New York: Vintage, 1973), ch. 9; and J.L. Mackie, *Ethics: Inventing Right and Wrong* (Oxford: Oxford Univ. Press, 1977).

[10] See. for example, K. Koslicki, *The Structure of Objects* (Oxford: Oxford Univ. Press, 2007).

[11] See Hanna, *Deep Freedom and Real Persons*, esp. chs. 6-7.

[12] See, for example, L. Alexander and M. Moore, "Deontological Ethics," *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, E.N. Zalta (ed.), available online at URL = <http://plato.stanford.edu/archives/fall2008/entries/ethics-deontological/>.

[13] W.D. Ross, *The Right and the Good* (Oxford: Oxford Univ. Press, 1930/2002), p. 22.

[14] See, for example, S. Scheffler, *Human Morality* (Oxford: Oxford Univ. Press, 1993).

[15] Ross, *The Right and the Good*, p. 158.

[16] Parfit, *On What Matters*, vol. 1, p. xliv.

[17] See R. Hanna, *Cognition, Content, and the A Priori: A Study in the Philosophy of Mind and Knowledge* (Oxford: Oxford Univ. Press, 2015), section 4.7, and chs. 6-8.

[18] See R. Hanna, "If God's Existence is Unprovable, Then is Everything Permitted? Kant, Radical Agnosticism, and Morality," DIAMETROS 39 (2014): 26-69; R. Hanna, "Radical Enlightenment: Existential Kantian Cosmopolitan Anarchism, With a Concluding Quasi-Federalist Postscipt," in D. Heidemann and K. Stoppenbrink (eds.), *Join, Or Die: Philosophical Foundations of Federalism* (Berlin: De Gruyter, 2016), pp. 63-90; Hanna, *Kant, Agnosticism, and Anarchism*; and Hanna, "Exiting the State and Debunking the State of Nature," THE RATIONAL HUMAN CONDITION, Vol. 1, essay 2.1.

[19] See Hanna, *Deep Freedom and Real Persons*, ch. 4.

[20] As I pointed out in n. 6 above, it's interesting that Shafer-Landau's *Fundamentals of Ethics* deviates from this all-too-familiar tripartite division, and, in its first two divisions, mirrors the classical post-Kantian *Sittlichkeit* vs. *Moralität* distinction.

[21] See S. McKeever and M. Ridge, (eds.), *Principled Ethics* (Oxford: Oxford Univ. Press, 2006).

[22] See Hanna, *Deep Freedom and Real Persons*, ch. 2.

[23] See note 5 above.

[24] I don't know what specifically motivates Kant's metaphysical caution in the *Groundwork*, nor do I know why he so overstates the formalism and rigorism of his moral theory there. These are ironic facts, and a real pity. The *Groundwork* is the most accessible and widely-read of all Kant's books, and it is also the book most closely studied by recent and contemporary Kantian ethicists. Yet in some ways it is *the least Kantian* of all Kant's books.

[25] I've elided "though only practical" from this Kant-text because it gives the false rhetorical impression that the *practical reality* of free will is somehow *less* metaphysically real than the *theoretical reality* of free will would be. On the contrary, free will is *fully metaphysically real*, and the "practical reality" of the fact of free will, as opposed to its "theoretical reality," means only that free will is *non-conceptually* veridically accessed at the source of agency, as opposed to its being *conceptually and propositionally* accessed there.

[26] See Hanna, *Deep Freedom and Real Persons*, section 2.3.

[27] See, for example, R. Hanna, "Kant, Causation, and Freedom," *Canadian Journal of Philosophy* 36 (2006): 281-306; R. Hanna, *Kant, Science, and Human Nature* (Oxford: Oxford Univ. Press, 2006), ch. 8; R. Hanna, "Freedom, Teleology, and Rational Causation," *Kant Yearbook* 1 (2009): 99-142; and Hanna, *Deep Freedom and Real Persons*, esp. ch. 3.

[28] See *In a Lonely Place* (directed by N. Ray, 1950).

[29] See, for example, S. Harris, *Free Will* (New York: Free Press, 2012).

[30] See J. Schuessler, "Philosophy That Stirs the Waters," *New York Times* (29 April 2013), available online at URL = <http://www.nytimes.com/2013/04/30/books/daniel-dennett-author-of-intuition-pumps-and-other-tools-for-thinking.html?emc=eta1&_r=0>.

[31] See T. Huxley, "On the Hypothesis That Animals are Automata, and Its History," in D. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings* (New York: Oxford Univ. Press, 2002), pp. 24-30, at pp. 29-30.

[32] See Hanna, *Deep Freedom and Real Persons*, chs. 3-5; and H. Steward, *A Metaphysics for Freedom* (Oxford: Oxford Univ. Press, 2012).

[33] See, for example, Korsgaard, *Self-Constitution: Agency, Identity, and Integrity*, p. 39; and Parfit, *On What Matters*, vol. 1, ch. 11.

[34] See Hanna, *Deep Freedom and Real Persons*, sections 4.5 and 7.2.

[35] See, for example, J. Rachels, *The Elements of Moral Philosophy* (5th edn., New York: McGraw-Hill, 2007), ch. 2; Williams, *Ethics and the Limits of Philosophy*, ch. 9; and B. Williams, *Moral Luck* (Cambridge: Cambridge Univ. Press, 1981), ch. 11.

[36] See, for example, Mackie, *Ethics: Inventing Right and Wrong*; R. Joyce, *The Myth of Morality* (Cambridge: Cambridge Univ. Press, 2001); and R. Joyce, *The Evolution of Morality* (Cambridge: MIT Press, 2006).

[37] See, for example, J. Dancy, *Ethics Without Principles* (Oxford: Oxford Univ. Press, 2004); J. Dancy, "Moral Particularism," *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, E.N. Zalta (ed.), URL = http://plato.stanford.edu/archives/fall2008/entries/moral-particularism/.

[38] See, for example, W.G. Sumner, *Folkways* (Boston: Ginn, 1906).

[39] Ross, *The Right and the Good*, ch. II.

[40] See note 37 above, and also J. Dancy, *Practical Reality* (Oxford: Oxford Univ. Press, 2000).

[41] See R. Hanna, *Rationality and Logic* (Cambridge: MIT Press, 2006), ch. 7; and Hanna, *Cognition, Content, and the A Priori*, ch. 5.

[42] Of course, it is an open question whether any real-life, real-world professional academic philosopher *ever* gives up a strongly-held view merely because there are sound philosophical arguments against it. At the same time, however, professional academic philosophers will *often enough* give up or suppress a strongly-held view if their professional academic status—for example, getting a job, receiving tenure and promotion, receiving regular raises, or receiving grants and fellowships—depends on it. That is a very depressing pair of facts.

[43] See, for example, J. Meiland and M. Krausz (eds.), *Relativism* (Notre Dame, IN: Univ. of Notre Dame Press, 1982); and also B. Hooker and M. Little (eds.), *Moral Particularism* (Oxford: Clarendon/Oxford Univ. Press, 2000).

[44] The various relationships between Existential Kantian Ethics and Aristotelian virtue ethics, Humean ethics, contractualism, and rule consequentialism are all more subtle, and not necessarily oppositional. Indeed, it is plausibly arguable that all five can be unified within a single suitably refined Kantian ethical framework. On the Aristotle-Hume-Kant connection, see C. Korsgaard, *The Constitution of Agency* (Oxford: Oxford Univ. Press, 2008); Korsgaard, *Creating the Kingdom of Ends*, chs. 8-11; and Korsgaard, *Self-Constitution: Agency, Identity, and Integrity*. And on the contractualism-rule consequentialism-Kantian ethics connection, see Parfit, *On What Matters*, vol. 1, esp. ch. 17.

[45] See Hanna, *Deep Freedom and Real Persons*, esp. chs. 3-7.

[46] See Hanna, *Cognition, Content, and the A Priori*, esp. chs. 2-3 and 5-8.

[47] This section draws directly, in places, on Hanna, *Rationality and Logic*, section 1.3.

[48] J.W. Goethe, *Faust*, line 2129: "Im Anfang war die Tat."

[49] L. Wittgenstein, Tractatus Logico-Philosophicus, trans. C.K. Ogden (London: Routledge & Kegan Paul, 1981), prop. 6.43, p. 185.

[50] See, for example, J. Lenman, "Moral Naturalism," *The Stanford Encyclopedia of Philosophy (Winter 2008 Edition)*,

E.N. Zalta (ed.), available online at URL = <http://plato.stanford.edu/entries/naturalism-moral/>.

[51] Moore, *Principia Ethica*, p. 40.

[52] Moore, *Principia Ethica*, p. 58.

[53] Moore, *Principia Ethica*, p. 73.

[54] Moore fails to distinguish between *concepts* and *properties*, and also between *properties* and *predicates*. See, for example, H. Putnam, "On Properties," in H. Putnam, *Mathematics, Matter, and Method: Philosophical Papers, Volume 1* (2nd edn., Cambridge: Cambridge Univ. Press, 1979), pp. 305-322; G. Bealer, *Quality and Concept* (Oxford: Clarendon/Oxford Univ. Press, 1982); and A. Oliver, "The Metaphysics of Properties," *Mind* 105 (1996): 1-80. This is, of course, controversial territory. But for my purposes I will make the fairly standard assumptions that concepts are intersubjectively-accessible psychological intensional entities whose identity criterion is definitional equivalence; that predicates are linguistic intensional entities whose identity criterion is synonymy; and that properties are non-psychological, non-linguistic intensional entities whose identity criterion is sharing cross-possible-worlds extensions. Predicates express concepts as their meanings, and concepts pick out corresponding properties in the world. For convenience, however, in the following discussion of Moore's failed argument against Ethical Naturalism I will allow 'property' to range over all three sorts of intensional entity. The flaws in his argument will persist no matter which sort of intensional entity is at issue.

[55] Moore, *Principia Ethica*, p. 15.

[56] Moore, *Principia Ethica*, p. 44.

[57] Indeed, Moore adopts Bishop Butler's dictum, "everything is what it is and not another thing," as the motto of *Principia Ethica*, and also uses it repeatedly as an axiom in his arguments.

[58] Moore, *Principia Ethica*, pp. 16-17.

[59] It is very likely that Moore inherited the phenomenological criterion of the identity of properties from his teacher James Ward, who in turn inherited it from Franz Brentano—thus by an ironic twist returning us full-circle to Naturalism, via Psychologism. See J. Ward, "Psychology," *Encyclopedia Brittanica* (29 vols., 11th edn., New York: Encyclopedia Brittanica, 1911), vol. 22, pp. 547-604; and Hanna, *Rationality and Logic*, ch. 1.

[60] Moreover the phenomenological criterion of property identity leads directly to The Paradox of Analysis: If (i) only *phenomenological* identity will suffice for property identity, and if (ii) property identity is a necessary condition of a correct analysis, then (iii) every correct analysis must be epistemically trivial. See C.H. Langford, "The Notion of Analysis in Moore's Philosophy," in P. Schilpp (ed.), *The Philosophy of G.E. Moore* (New York: Tudor, 1952), pp. 321-342; and also Moore's reply to Langford, "Analysis," in Schilpp (ed.), *The Philosophy of G.E. Moore*, pp. 660-667.

[61] Moore, *Principia Ethica*, pp. 206-207.

[62] G.E. Moore, "The Conception of Intrinsic Value," in G.E. Moore, *Philosophical Studies* (New York: Harcourt Brace, 1922), pp. 253-275, at p. 261.

[63] See Hanna, *Cognition, Content, and the A Priori*, ch. 5.

[64] See G. Bealer, "Modal Epistemology and the Rationalist Renaissance," in T. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility* (Oxford: Oxford Univ. Press, 2002), pp. 71-125.

[65] R. Hanna and M. Maiese, *Embodied Minds in Action* (Oxford: Oxford Univ. Press, 2009), esp. chs. 3-5.

[66] See H. Frankfurt, "Freedom of the Will and the Concept of a Person," "Identification and Externality," "Identification and Wholeheartedness," and "The Importance of What We Care About," all in H. Frankfurt, *The Importance of What We Care About* (Cambridge: Cambridge Univ. Press, 1988), pp. 11-25, 58-68, 159-176, and 80-94.

[67] See Frankfurt, "The Problem of Action," in Frankfurt, *The Importance of What We Care About*, pp. 42-52.

[68] B. O'Shaughnessy, "Trying (as the Mental 'Pineal Gland')," *Journal of Philosophy* 70 (1973): 365-386.

[69] R. Cummins, "Reflections on Reflective Equilibrium," in M.R. DePaul and W. Ramsey (eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry* (Lanham, MD: Rowman and Littlefield, 1998), pp. 113-127, at p. 125.

[70] See R. Hanna, *Kant and the Foundations of Analytic Philosophy* (Oxford: Oxford Univ. Press, 2001), ch. 3; and Hanna, *Cognition, Content, and the A Priori*, ch. 4.

[71] See Hanna, *Rationality and Logic*, ch. 1; see also R. Hanna, "Husserl's Arguments against Logical Psychologism," in V. Mayer (ed.), *Husserls Logische Untersuchungen* (Munich: Akademie Verlag, 2008), pp. 27-42.

[72] See Hanna, "A Kantian Critique of Scientific Essentialism," *Philosophy and Phenomenological Research* 58 (1998): 497-528; Hanna, *Kant, Science, and Human Nature*, ch. 3; and Hanna, *Cognition, Content, and the A Priori*, section 4.5.

[73] D. Barnett, "Against A Posteriori Moral Naturalism," *Philosophical Studies* 107 (2002): 239-257.

[74] See Hanna, *Cognition, Content, and the A Priori*, chs. 2, 4, and 5.

[75] See Hanna, *Cognition, Content, and the A Priori*, section 4.7.

[76] See Hanna, *Cognition, Content, and the A Priori*, ch. 8.

[77] See Hanna, *Rationality and Logic*, chs. 5-6.

[78] See Hanna, *Cognition, Content, and the A Priori*, chs. 6 to 8.

[79] More generally, see A. Chapman, A. Ellis, R. Hanna, T. Hilderbrand, and H. Pickford, *In Defense of Intuitions: A New Rationalist Manifesto* (London: Palgrave Macmillan, 2013).

[80] See Hanna and Maiese, *Embodied Minds in Action*, chs. 3-5.
[81] By which I mean: as long as you are still a sentient, conscious, desiring, caring, self-conscious, and minimally rational living organism.
[82] See H. Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil* (Harmondsworth, Middlesex: Penguin, 1977); see also Hanna, *Deep Freedom and Real Persons*, section 3.3.
[83] D. Richie, *Ozu* (Berkeley and Los Angeles, CA: Univ. of California Press, 1974), p. 252.
[84] R.W. Emerson, "Self-Reliance," in S.E. Whicher (ed.), *Selections from Ralph Waldo Emerson* (Boston, MA: Houghton Mifflin, 1957), pp. 147-168, at 153.
[85] Ross, *The Right and the Good*, pp. 29-30.
[86] D. Gahan, M. Gore, and A. Fletcher, "Policy of Truth," *Violator* (1990).
[87] J.-P. Sartre, "Existentialism is a Humanism," trans. B. Frechtman, in S. Cahn and P. Markie (eds.), *Ethics: History, Theory, and Contemporary Issues* (3rd edn., New York: Oxford Univ. Press, 2006), pp. 396-402, at 400.
[88] A. Ahuja, "An Organized Death," *London Times* (4 September 2000), available online at URL = < http://www.cavehill.uwi.edu/BNCCde/_e%26ae/times_features.htm>. See also section 6.9 below.
[89] See also S. Baiasu, *Kant and Sartre: Re-Discovering Critical Ethics* (London: Palgrave Macmillan, 2011).
[90] See, for example, O. O'Neill, "Consistency in Action," in Cahn and Markie (eds.), *Ethics: History, Theory, and Contemporary Issues*, pp. 541-558.
[91] See, for example, C. Korsgaard, "The Right to Lie: Kant on Dealing with Evil," in Korsgaard, *Creating the Kingdom of Ends*, pp. 133-158. The flip side of the rigorism problem is that what Kant variously calls "meritorious duties," "imperfect duties," or "duties of virtue" do not seem to be strict or universal *enough*. One might call this the *under-rigorism* problem. In section 2.2 below, I will offer a solution to the rigorism problem that is also intended to handle the under-rigorism problem.
[92] See, for example, T. Hill, "Moral Dilemmas, Gaps, and Residues," in T. Hill, *Human Welfare and Moral Worth: Kantian Perspectives* (Oxford: Clarendon/Oxford University Press, 2002), essay 12, pp. 362-402; and H.E. Mason (ed.), *Moral Dilemmas and Moral Theory* (Oxford: Oxford Univ. Press, 1996).
[93] See, for example, G. Boolos and R. Jeffrey, *Computability and Logic* (3rd edn.; Cambridge: Cambridge Univ. Press, 1989), chs. 10, 22, and 25, and esp. pp. 250-255.
[94] See Hanna, *Cognition, Content, and the A Priori*, ch. 5.
[95] See Hanna, *Cognition, Content, and the A Priori*, ch. 4.
[96] See Hanna, *Kant and the Foundations of Analytic Philosophy*; and Hanna, *Cognition, Content, and the A Priori*, esp. chs. 2, 4, and 5. . See also R. Hanna, "Kant's Theory of Judgment," *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), E.N. Zalta (ed.), available online at URL = <https://plato.stanford.edu/archives/win2017/entries/kant-judgment/>.
[97] R. Stern, "Does 'Ought' Imply 'Can'? And Did Kant Think It Does?" *Utilitas* 16 (2004): 42-61, at p. 59.
[98] Ross, *The Right and the Good*, p. 21.
[99] Ross, *The Right and the Good*, pp. 20, 33, and 41.
[100] Ross, *The Right and the Good*, pp. 30-31, and 145-148.
[101] Hanna, *Cognition, Content, and the A Priori*, chs. 6-8.
[102] See Mackie, *Ethics: Inventing Right and Wrong*. See also section 1.2 above.
[103] J. Rawls, *A Theory of Justice* (Cambridge, MA: Harvard Univ. Press, 1971), p. 41. Rawls correctly points out that unless Ross explicitly works out a monistic objective deontological lexical ordering or weighting scale for his prima facie duties in relation to actual act-contexts, then his view is a dead end. My objection is just the flip side of this, namely that Ross is clearly implicitly presupposing a monistic objective deontological lexical ordering or weighting scale for mapping his prima facie judgments to actual duties in actual contexts, without either admitting it or telling us precisely what it is. Many thanks to Eric Lee for pointing this out to me, and also to Kevin White for drawing my attention to the Rawls parallel.
[104] See Hanna, *Cognition, Content, and the A Priori*, chs. 6-8.
[105] See, for example, P. Benacerraf, "What Numbers Could Not Be," *Philosophical Review* 74 (1965): 47-73; S. Shapiro, *Philosophy of Mathematics: Structure and Ontology* (New York: Oxford Univ Press, 1997); and S. Shapiro, *Thinking about Mathematics* (Oxford: Oxford Univ. Press, 2000), ch. 10.
[106] See C. Parsons, *Mathematical Thought and its Objects* (Cambridge: Cambridge Univ. Press, 2008), esp. chs. 3, 5-6, and 9.
[107] See, for example, Hanna, *Kant and the Foundations of Analytic Philosophy*, chs. 1-2; Hanna, *Kant, Science, and Human Nature*, ch. 6; and R. Hanna, *Rationality and Logic*, chs. 4 and 6.
[108] See, for example, T. Hill, "Kantian Constructivism in Ethics," *Ethics* 99 (1989): 752-770; and O'Neill, *Constructions of Reason*.
[109] See Hanna, *Cognition, Content, and the A Priori*, section 4.6.
[110] See Hanna, *Rationality and Logic*, ch. 7; and Hanna, *Cognition, Content and the A Priori*, ch. 5. See also R. Hanna, "Rationality and the Ethics of Logic," *Journal of Philosophy* 103 (2006): 67-100.
[111] See, for example, G. Priest, *In Contradiction* (Dordrecht: Martinus Nijhoff, 1987); and G. Priest, "What is So Bad About Contradictions?," *Journal of Philosophy* (1998): 410-426.
[112] See, for example, A. Tarski, "The Semantic Conception of Truth and the Foundations of Semantics," *Philosophy*

and Phenomenological Research 4 (1943-44): 341-375.

[113] See, for example, K. Gödel, "On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems," in J. Van Heijenoort (ed.), *From Frege to Gödel* (Cambridge, MA: Harvard Univ. Press, 1967), pp. 596-617.

[114] H. Putnam, "There is At Least One A Priori Truth," in H. Putnam, *Realism and Reason: Philosophical Papers, Vol. 3* (Cambridge: Cambridge Univ. Press, 1983), pp. 98-114, at pp. 100-101 (italics in the original).

[115] For details, see Hanna, *Cognition, Content, and the A Priori*, ch. 7.

[116] See, for example, Hanna, *Kant and the Foundations of Analytic Philosophy*, section 5.1; Hanna, *Deep Freedom and Real Persons*, chs. 1-5; and Hanna, *Cognition, Content, and the A Priori*, sections 3.3, and 4.7.

[117] See also Hanna, *Kant, Science, and Human Nature*, esp. ch. 8; Hanna, "Freedom, Teleology, and Rational Causation"; and Steward, *A Metaphysics for Freedom*.

[118] See Hanna and Maiese, *Embodied Minds in Action*.

[119] See, for example, B. Williams, "Ethical Consistency," in B. Williams, *Problems of the Self* (Cambridge: Cambridge Univ. Press, 1973), pp. 166-186; B. Williams, "Moral Luck" and "Conflicts of Values," both in B. Williams, *Moral Luck* (Cambridge: Cambridge Univ. Press, 1981), pp. 20-39 and 71-82; and M. Nussbaum, *The Fragility of Goodness* (Cambridge: Cambridge Univ. Press, 1986).

[120] See, for example, Hill, "Moral Dilemmas, Gaps, and Residues"; A. Donagan, "Consistency in Rationalist Moral Systems," *Journal of Philosophy* 81 (1984): 291-309; and A. Donagan, "Moral Dilemmas, Genuine and Spurious: A Comparative Anatomy," *Ethics* 104 (1993): 7-21.

[121] See, for example, G. Sayre-McCord, "A Moral Argument Against Moral Dilemmas," available online at URL= <http://philosophy.unc.edu/people/faculty/geoffrey-sayre-mccord/on-line-apers/A%20Moral%20Argument%20Against%20Moral%20Dilemmas.pdf>.

[122] See, for example, the papers or books cited in notes 120-122 above. See also S. Shiffrin, *Speech Matters: On Lying, Morality, and the Law* (Princeton, NJ: Princeton Univ. Press, 2016), esp. ch. 1.

[123] See T. Schapiro, "Kantian Rigorism and Mitigating Circumstances," *Ethics* 117 (2006): 32-57.

[124] See also Hanna, *Cognition, Content, and the A Priori*, section 4.5.

[125] See also, for example, D. Jacquette, "Moral Dilemmas, Disjunctive Obligations, and Kant's Principle That 'Ought' implies 'Can'," *Synthese* 88 (1991): 43-55.

[126] See, for example, M. Unamuno, *The Tragic Sense of Life*, available online at URL = <http://www.gutenberg.org/files/14636/14636-h/14636-h.htm>.

[127] On moral nihilism, see also A. Camus, *The Rebel*, trans. A. Bower (New York: Vintage, 1956).

[128] See A. Tarski, "The Semantic Conception of Truth and the Foundations of Semantics," *Philosophy and Phenomenological Research* 4 (1943-44): 341-375.

[129] S. Kripke, "Outline of a Theory of Truth," *Journal of Philosophy* 72 (1975): 690-715.

[130] R.B. Marcus, "Moral Dilemmas and Consistency," *Journal of Philosophy* 77 (1980): 121-136.

[131] On the important distinction between *deep* moral responsibility and *shallow* moral responsibility, see Hanna, *Deep Freedom and Real Persons*, esp. chs. 4-5.

[132] Directed by Michael Curtiz (1942).

[133] The force of this analogy between Sartre's case and the *Casablanca* case of course depends to some extent on how one thinks about mother-son relationships vs. romantic love-relationships vs. marriage-relationships (in short, how we think about the Anna Karenina scenario), not to mention how we fill in various important details that are not provided either by Sartre or by the many scriptwriters of *Casablanca*. But to make my point more intuitively vivid, imagine that Laszlo had been played by a relatively unattractive actor with a ponderous manner—say, Basil Rathbone or Ralph Richardson, who played Karenin in the 1935 and 1948 film versions of *Anna Karenina* respectively—and not by the more attractive and charismatic Paul Henreid.

[134] D. Marquis, "Why Abortion is Immoral," *Journal of Philosophy* 86 (1989): 183-202, at 202.

[135] M. Tooley, "Abortion and Infanticide," in M. Cohen, T. Nagel, and T. Scanlon (eds.), *The Rights and Wrongs of Abortion* (Princeton, NJ: Princeton Univ. Press, 1974), pp. 52-84, at pp. 52-53.

[136] See also F. Kamm, *Morality, Mortality*, 2 vols. (New York: Oxford Univ. Press, 1993/1996); F. Kamm, *Intricate Ethics* (Oxford: Oxford Univ. Press, 2007); and J. McMahan, *The Ethics of Killing* (Oxford: Oxford Univ. Press, 2002).

[137] See, for example, the studies cited and discussed in D. Boonin, *A Defense of Abortion* (Cambridge: Cambridge Univ. Press, 2003), ch. 3. For an earlier study that puts the emergence of consciousness at between 22-26 weeks, see *British Parliamentary Office of Science and Technology Notes* 94 (1997), available online at URL = <http://www.parliament.uk/post/pn094.pdf>.The earlier study was done by Prof. Maria Fitzgerald of the Dept. of Anatomy and Developmental Biology at University College London in 1995.

[138] See, for example, Boonin, *A Defense of Abortion*; F. Kamm, *Creation and Abortion* (Oxford: Oxford Univ. Press, 1992); R. Hursthouse, "Virtue Theory and Abortion," in S. Cahn and P. Markie (eds.), *Ethics: History, Theory, and Contemporary Issues* (3rd edn.; Oxford: Oxford Univ. Press, 2006), pp. 765-778; Marquis, "Why Aborton is Immoral"; J. McMahan, *The Ethics of Killing*, ch. 4; J.J. Thomson, "A Defense of Abortion," in Cohen, Nagel, and Scanlon (eds.), *The Rights and Wrongs of Abortion*, pp. 3-22; Tooley, "Abortion and Infanticide"; and M.A. Warren, "On the Moral and Legal Status of Abortion," *Monist* 57 (1973): 43-61.

[139] See Thomson, "A Defense of Abortion," p. 22.
[140] See also Hanna, *Deep Freedom and Real Persons*, section 6.0.
[141] The thesis of epigenesis in biology says that biological material is initially unformed and that form gradually emerges through the non-predetermined or relatively spontaneous operations of an innate endogenous organizational or processing device in interaction with its environment. See, for example, J. Maienschein, "Epigenesis and Preformationism," *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/epigenesis/>; Hanna, *Deep Freedom and Real Persons*, ch. 1; and Hanna, "Kant, Nature, and Humanity," THE RATIONAL HUMAN CONDITION Vol. 1, essay 2.2.
[142] See Hanna and Maiese, *Embodied Minds in Action*, esp. chs. 1-5; Hanna, *Deep Freedom and Real Persons*,esp. chs. 1-5; and also R. Hanna, "Minding the Body," *Philosophical Topics* 39 (2011): 15-40.
[143] See also Hanna, *Deep Freedom and Real Persons*, ch. 2.
[144] For more on dynamic emergence, see Hanna and Maiese, *Embodied Minds in Action*, ch. 8.
[145] See, for example, Hanna, *Rationality and Logic*, esp. chs. 1, 4, and 5.
[146] See ch. 6 below for further justification of that claim.
[147] That is: we ought to treat people in ways that are sufficient to meet the demands of respect for their human dignity. This is a specifically Kantian version of what Harry Frankfurt calls "the doctrine of sufficiency"; see his *On Inequality* (Princeton, NJ: Princeton Univ. Press, 2015).
[148] —As opposed to her merely *self-perceived and false human needs*, that is. It might be that someone perceives within herself an intense need to own a certain luxury automobile, even though she already owns a car that is perfectly adequate to her true human needs. Therefore it is not unkind of us not to cater to this self-perceived and false human need. For more on the crucial distinction between true human needs and false human needs, see M. Maiese and R. Hanna, *The Mind-Body Politic* (London: Palgrave Macmillan, forthcoming in 2019), ch. 3.
[149] See, for example, BBC News, "Man Held for German 'Cannibal Killing'," available online at URL = <http://news.bbc.co.uk/2/hi/europe/2569095.stm>.
[150] J.S. Mill, *On Liberty* (Indianapolis, IN: Hackett, 1978), pp. 101.
[151] See Hanna, *Deep Freedom and Real Persons*, chs. 1 to 5; and Hanna, "Kant, Nature, and Humanity," THE RATIONAL HUMAN CONDITION, Vol. 1, essay 2.2.
[152] See, for example, D. Pereboom, *Living Without Free Will* (Cambridge: Cambridge Univ. Press, 2001); Steward, *A Metaphysics for Freedom*; and Hanna, *Deep Freedom and Real Persons*, sections 4.5 and 7.2.
[153] See Hanna, *Kant, Agnosticism, and Anarchism*, sections 18-20.
[154] See esp. sections 6.7 to 6.9 below.
[155] When I was a graduate student at Yale in the 1980s, John Smith reported to some of us that J.N. Findlay had used the phrase "associate membership in the Kingdom of Ends" in lectures at Yale. I do not know precisely what Findlay himself meant by this phrase: but it is so appropriate that I cannot resist stealing it and adapting it for my own philosophical purposes.
[156] Strictly speaking, there is also a morally and politically important distinction between (i) *primary* coercion and (ii) *secondary* coercion, such that in primary coercion, the coercer uses *violence or threats of violence*, whereas in secondary coercion, the coercer uses only salient harm or threats of salient harm *that fall short of violence*—e.g., unofficially slandering you, name-shaming you over the internet, "doxxing" you, officially reprimanding you, officially fining you, firing you from your job, blacklisting you, etc. See, e.g., Hanna, *Kant, Agnosticism, and Anarchism*, part 2. But introducing that distinction here would needlessly complicate my exposition in this argument-context.
[157] Thomson, "A Defense of Abortion," p. 738.
[158] J.J. Thomson, "Self-Defense," *Philosophy and Public Affairs* 20 (1991): 283-310; Thomson's own term is "innocent threat."
[159] See Rawls, *A Theory of Justice*.
[160] See section 6.8 below.
[161] See, for example, B. Williams, "A Critique of Utilitarianism," in J.J.C. Smart and B. Williams, *Utilitarianism: For and Against* (Cambridge: Cambridge Univ. Press, 1973), pp. 77-50.
[162] See, for example, D. Lewis, *Convention* (Cambridge: Harvard Univ. Press, 1969); and
M. Rescorla, "Convention," *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), E.N. Zalta (ed.), available online at URL = <https://plato.stanford.edu/archives/sum2017/entries/convention/>.
[163] See, for example, L. Wenar, "Rights," *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), E.N. Zalta (ed.), available online at URL = <https://plato.stanford.edu/archives/fall2015/entries/rights/>;
and J.J. Thomson, *The Realm of Rights* (Cambridge: Harvard Univ. Press, 1990).
[164] See Hanna, *Kant, Agnosticism, and Anarchism*, section 3.14.
[165] See Boonin, *A Defense of Abortion*, sections 2.1-2.5; J. Finnis, "The Rights and Wrongs of Abortion," in Cohen, Nagel, and Scanlon (eds.), *The Rights and Wrongs of Abortion*, pp. 85-113; and J.J. Thomson, "Rights and Deaths," in Cohen, Nagel, and Scanlon (eds.), *The Rights and Wrongs of Abortion*, pp. 114-127.
[166] See Boonin, *A Defense of Abortion*, sections 2.6-2.8.

[167] See Tooley, "Abortion and Infanticide"; and Warren, "On the Moral and Legal Status of Abortion."
[168] See Thomson, "A Defense of Abortion"; and Boonin, *A Defense of Abortion*, ch. 4.
[169] See Kamm, *Creation and Abortion*; and Boonin, *A Defense of Abortion*.
[170] Many thanks to Onora O'Neill for stressing this point to me, in conversation.
[171] See also O. Sensen, *Kant on Human Dignity* (Berlin/Boston: Walter de Gruyter, 2011).
[172] In such cases, it remains possible, according to associate membership in the Realm of Ends, for a *convention-based* right-to-life to be extended to these non-persons. But the fact of such exceptions is not a consequence of the Low Bar of Personhood approach, as such, and therefore its existence does not count as support for that approach.
[173] Directed by Robert Wise (1951).
[174] See D. Adams, *The Hitchhiker's Guide to the Galaxy* (New York: Ballantine/Del Rey, 2002).
[175] See S. Lem, *Fiasco*, trans. M. Kandel (New York: Harcourt Brace Jovanovich, 1987).
[176] See also H. Kuhse and P. Singer, "Individuals, Humans, and Persons: The Issue of Moral Status," in P. Singer, H. Kuhse, S. Buckle, K. Dawson, and P. Kasimba (eds.), *Embryo Experimentation* (Cambridge: Cambridge Univ. Press, 1990), pp. 65-75; and P. Singer, *Practical Ethics* (Cambridge: Cambridge Univ. Press, 1993), pp. 152-156.
[177] See, for example, Hanna, *Cognition, Content, and the A Priori*, section 3.3.
[178] See also, for example, M. Tomasello, *Why We Cooperate* (Cambridge, MA: MIT Press, 2009); and P. Bloom, "The Moral Life of Babies," *New York Times Magazine* (9 May 2010).
[179] See J. Rawls, "The Independence of Moral Theory," *Proceedings and Addresses of the American Philosophical Association* 48 (1974): 5-22; J. Rawls, *A Theory of Justice* (Cambridge, MA: Harvard Univ. Press, 1971); and N. Daniels, "Wide Reflective Equilibrium and Theory Acceptance in Ethics," *Journal of Philosophy* 76 (1979): 256-282.
[180] Thomson, "A Defense of Abortion," p. 13.
[181] See, for example, P. Singer, "Famine, Affluence, and Morality,"in Cahn and Markie (eds.), *Ethics: History, Theory, and Contemporary Issues*, pp. 789-796; Singer, *Practical Ethics*, ch. 8; and P. Singer, "The Singer Solution to World Poverty," in J. Rachels and S. Rachels, (eds.), *The Right Thing to Do: Basic Readings in Moral Philosophy* (4th edn., New York: McGraw Hill, 2007), pp. 138-144.
[182] See note 176 above.
[183] Tooley, "Abortion and Infanticide," pp. 73-77.
[184] See T. Nagel, "What Is It Like To Be A Bat?," in T. Nagel, *Mortal Questions* (Cambridge: Cambridge Univ. Press, 1979), pp. 165-180.
[185] D. Dennett, "Conditions for Personhood," in D. Dennett, *Brainstorms* (Cambridge, MA: MIT Press, 1976), pp. 267-285; and M.A. Warren, *Moral Status* (Oxford: Oxford Univ. Press, 2000).
[186] See Hanna, *Cognition, Content, and the A Priori*, section 4.3.
[187] See, for example, J. Finnis, *Natural Law and Natural Rights* (Oxford: Clarendon/Oxford Univ. Press, 1980).
[188] Nagel, "What Is It Like To Be A Bat?," pp. 169-170.
[189] J. Bentham, *Principles of Morals and Legislation* (New York: Hafner, 1948), p. 311.
[190] P. Singer, "All Animals are Equal," in P. Singer, *Unsanctifying Human Life* (Oxford: Blackwell, 2002), pp. 80-94, at p. 84.
[191] See also Hanna, *Deep Freedom and Real Persons*, chs. 6 to 8.
[192] See, for example, P. Carruthers, *The Animals Issue* (Cambridge: Cambridge Univ. Press, 1992).
[193] See, for example, Singer, "All Animals are Equal."
[194] See T. Regan, *The Case for Animal Rights* (Berkeley, CA: Univ. of California Press, 1983); Singer, "All Animals are Equal"; and Singer, *Practical Ethics*, ch. 5.
[195] Actually this needs refinement in the case of Singer, whose considered view (often not known, or noted, or accepted, by Singer's own supporters) is in fact that although all sentient animals deserve equality of moral consideration with persons from the initial standpoint of act utilitarian deliberation, nevertheless *persons* lexically rank more highly than non-persons in calculating overall utility. See P. Singer, "Killing Humans and Killing Animals" in Singer, *Unsanctifying Human Life*, pp. 112-122. This in turn enables Singer to avoid an absurd answer to the notorious "lifeboat" puzzle, which runs as follows: If there is one human person and one large dog in a lifeboat far out at sea, and only enough food for one of the two creatures to survive, which creature goes into the water? Regan's view implausibly entails that the human person should flip a coin, and sacrifice herself if she loses. By contrast, Singer's view plausibly entails that the dog should go. The Concern for All Minded Animals Theory converges with Singer's view here.
[196] See, for example, C. Allen and M. Bekoff, *Species of Mind* (Cambridge: MIT Press, 1997); M. Bearzi and C. Stanford, *Beautiful Minds: The Parallel Lives of Great Apes and Dolphins* (Cambridge: Harvard Univ. Press, 2008); D.R. Griffin, *Animal Minds* (Chicago: Univ. of Chicago Press, 2001); D.R. Griffin, *Animal Thinking* (Cambridge: Harvard Univ. Press, 1984); D.R. Griffin, *The Question of Animal Awareness* (New York: Rockefeller Univ. Press, 1976); and S. Savage-Rumbaugh and R. Lewin, *Kanzi: The Ape at the Brink of the Human Mind* (New York: Wiley, 1994).
[197] See Hanna, *Deep Freedom and Real Persons*, section 1.0.

[198] See Hanna, *Deep Freedom and Real Persons*, chs. 3 to 5.
[199] Directed by M.C. Cooper and E.B. Schoedsack (1933).
[200] Directed by D. Myrick and E. Sánchez (1999).
[201] J.M. Hawkins and R. Allen (eds.), *Oxford Encyclopedic English Dictionary* (Oxford: Clarendon Press/Oxford Univ. Press, 1991), p. 52.
[202] It is also true that necessarily, every real person has one and only one living animal body, and conversely, necessarily, every living animal body of a real person is lived by one and only one real person. See Hanna, *Deep Freedom and Real Persons*, section 6.2.
As of 2018, new biomedical evidence suggests that all women who are capable of becoming pregnant are in fact *totipotent* and *chimeras*, in that their DNA changes when they become pregnant, fusing with the DNA of the zygote and fetus, so that their biological individuality is not fixed until they have become either pregnant or else incapable of becoming pregnant. See K. Rowland, "We Are Multitudes," *Aeon* (11 January 2018), available online at URL = <https://aeon.co/essays/microchimerism-how-pregnancy-changes-the-mothers-very-dna>. If that is correct, then many or even most women do not have a unique living animal body until several decades after they are already real persons. This is a serious problem for Standard Animalism, which identifies people with individual living animal bodies, since it would then follow that many or even most women are not people for much of their lives—which is clearly absurd. But it is not a problem for Minded Animalism, which identifies people with each and all stages of their minded animal lives.
[203] See D. DeGrazia, *Taking Animals Seriously* (Cambridge: Cambridge Univ. Press, 1996); D. Dennett, "Animal Consciousness: What Matters and Why," in D. Dennett, *Brainchildren* (Cambridge: MIT Press, 1998), pp. 337-352; and D. Dennett, *Kinds of Minds: Toward an Understanding of Consciousness* (New York: Basic Books, 1996).
[204] For the "natural zombie" view, see S. Allen-Hermanson, "Insects and the Problem of Simple Minds: Are Bees Natural Zombies?," *Journal of Philosophy* 105 (2008): 389-415.
[205] See also Steward, *A Metaphysics for Freedom*, ch. 4, where she explicitly argues that
spiders and earthworms can be (in my terminology) proto-agents.
[206] See, for example, Hanna and Maiese, *Embodied Minds in Action*, chs. 1-2; and Hanna, *Cognition, Content, and the A Priori*, ch. 2.
[207] See, for example, P. Godfrey-Smith, *Other Minds: The Octopus and the Evolution of Intelligent Life* (New York: Collins, 2017); and A. Srinivasan, "The Sucker, the Sucker!," *London Review of Books* 39 (September 2017): 23-25, available online at URL =
<https://www.lrb.co.uk/v39/n17/amia-srinivasan/the-sucker-the-sucker?utm_source=newsletter&utm_medium=email&utm_campaign=3917&utm_content=usca_subs>.
[208] See, for example, C. Allen, "Animal Pain," *Noûs* 38 (2004): 617-643.
[209] See S. Rosenberg, S.K. Marie, and S. Kliemann, "Congenital Insensitivity to Pain with Anhidrosis (hereditary sensory and autonomic neuropathy type IV)," *Pediatric Neurology* 11 (1994): 50-56. Anhidrosis is lack of sweating.
[210] See R. Melzack, "Pain," in R. Gregory (ed.), *Oxford Companion to the Mind* (Oxford: Oxford Univ. Press, 1987), pp. 574-575.
[211] This corresponds to the possibility of what David Lewis calls *Martian pain*. See D. Lewis, "Mad Pain and Martian Pain," in N. Block (ed.), *Readings in the Philosophy of Psychology* (2 vols., Cambridge: Harvard Univ. Press, 1980), vol. 1, pp. 216-222.
[212] This corresponds to the possibility of what Lewis calls *mad pain*. See Lewis, "Mad Pain and Martian Pain."
[213] See J. Kim, *Philosophy of Mind* (Boulder, CO: Westview Press, 1998), p. 38.
[214] See H. Putnam, "Brains and Behavior," in H. Putnam, *Mind, Language, and Reality: Philosophical Papers, Vol. 2* (Cambridge: Cambridge Univ. Press, 1975), pp. 325-341.
[215] On the notion of a body-image, see M. Merleau-Ponty, *Phenomenology of Perception*, trans. C. Smith (London: Routledge, 1962), pp. 98-147; and S. Gallagher, *How the Body Shapes the Mind* (Oxford: Clarendon Press, 2005), ch. 3.
[216] See J. Fodor, "Special Sciences, or the Disunity of Science as a Working Hypothesis," in Block (ed.), *Readings in the Philosophy of Psychology*, vol. 1, pp. 120-133.
[217] See section 6.1 below.
[218] See, for example, B. Williams, "The Self and the Future," *Philosophical Review* 79 (1970): 161-180; D. Parfit, "Personal Identity," *Philosophical Review* 80 (1971): 3-27; and D. Parfit, *Reasons and Persons* (Oxford: Oxford Univ. Press, 1984), chs. 10-13.
[219] See Hanna, *Deep Freedom and Real Persons*, sections 3.3-3.4.
[220] An excellent example of this sort of oppression is the USA's healthcare system.
[221] See Hanna, *Kant, Agnosticism, and Anarchism*, parts 2 and 3.
[222] In other words, other things being equal, such birth-mothers—provided that, after the fact, they do *not* regret their choice—do *not* suffer during childbirth. If someone did not regret her choice and yet still *said* she had suffered, it would be plausible to think she was either just using language loosely or committing the Bentham-Singer fallacy.

[223] See J. Levine, "Materialism and Qualia: The Explanatory Gap," *Pacific Philosophical Quarterly* 64 (1983): 354-361.

[224] See also T. Nagel, "Conceiving the Impossible and the Mind-Body Problem," *Philosophy* 73 (1998): 337-352; and T. Nagel, "The Psychophysical Nexus," in P. Boghossian and C. Peacocke (eds.), *New Essays on the A Priori* (Oxford: Clarendon/Oxford Univ. Press, 2000), 433-471.

[225] Nagel, "What Is It Like To Be A Bat?," p. 166.

[226] (Dir. M. Gordon, 1959). And yes, for better or worse, I am just that kind of cinephile.

[227] Nagel, "Panpsychism," in Nagel, *Mortal Questions*, pp. 181-195, at 191, underlining added.

[228] This point is vividly brought out by Stanislaw Lem in his brilliant sci-fi novel, *Fiasco*. It's about the tragic impossibility of rational human contact with truly alien real persons. The alien Quintans, who are indeed real persons, are a living network of "coarse, bloated mounds" on the surface of the planet Quinta. Naturally, they're radically misunderstood by a team of contact-seeking Earthmen, and ultimately blasted into smithereens by them, even despite the Earthmen's best intentions.

[229] Nagel, "What Is It Like To Be A Bat?," p. 172.

[230] For example, the classical type-physicalist mind-brain identity theory defended by Place and Smart explicitly rejects analytical concept identity and also explicitly asserts contingent property identity. See U.T. Place, "Is Consciousness a Brain Process?," *British Journal of Psychology* 47 (1956): 44-50; and J.J.C. Smart, "Sensations and Brain Processes," *Philosophical Review* 68 (1959): 141-156.

[231] For example, prior to 2005, Kim was a reductive physicalist who also asserted the existence of a Mental-Mental Gap. See J. Kim, "Multiple Realization and the Metaphysics of Reduction," in J. Kim, *Supervenience and Mind* (Cambridge: Cambridge Univ. Press, 1993), pp. 309-335, at 334. Nowadays, Kim is a non-reductivist about consciousness, although he remains a reductive physicalist about intentionality—see J. Kim, *Physicalism, Or Something Near Enough* (Princeton, NJ: Princeton Univ. Press, 2005), and also presumably still holds The Mental-Mental Gap Thesis.

[232] See, for example, Lewis, "Mad Pain and Martian Pain."

[233] See, for example, Kim, *Philosophy of Mind*, chs. 4-5.

[234] For the functionalist, these inputs and outputs may be, but do not have to be, *stimulus* inputs and *verifiable* outputs. So Functionalism comprehends Behaviorism, yet extends well beyond it.

[235] More precisely, a functional property is the property of having a first-order physical property with a certain causal-role specification. See Putnam, "On Properties," at pp. 313-315.

[236] See notes 233 and 235 above.

[237] Kim, "Multiple Realization and the Metaphysics of Reduction," p. 313.

[238] See, for example, Hanna and Maiese, *Embodied Minds in Action*, section 8.1.

[239] See Kim, "Multiple Realizability and the Metaphysics of Reduction."

[240] See Hanna, *Cognition, Content, and the A Priori*, ch. 2.

[241] For the distinction between body schema and body image, see Hanna and Maiese, *Embodied Minds in Action*, pp. 68-70; and Gallagher, *How the Body Shapes the Mind*, pp. 37-38.

[242] See E. Thompson, "Empathy and Consciousness," *Journal of Consciousness Studies* 8 (2001): 1-32.

[243] See, for example, M. Maiese, *Embodiment, Emotion, and Cognition* (London: Palgrave MacMillan, 2011).

[244] See, for example, V. Hearne, *Adam's Task: Calling Animals by Name* (New York: Knopf, 1986).

[245] I say "experienced" here and not the slightly more specific "subjectively experienced," so as to include the class of non-human non-person proto-sentient or simple minded animals within the scope of The Moral Comparison Thesis.

[246] See J. Russell and R. Hanna, "A Minimalist Approach to the Development of Episodic Memory," *Mind and Language* 27 (2012): 29-54.

[247] For a general discussion of theories of punishment, see, e.g., D. Boonin, *The Problem of Punishment* (Oxford: Oxford Univ. Press, 2008).

[248] What do I mean by that? The most I can say here is this. First, Kant's own retributivist theory of punishment, and correspondingly his views on capital punishment, are both deeply mistaken, because retributivism and capital punishment not only license, but indeed morally require, direct violations of respect for human dignity. Second, all punishment is coercive; and coercion is immoral because it inherently involves treating people as mere means or mere things in order to promote various ends of the coercer; hence *all punishment is immoral*, whether its purported justification is retributive, deterrent, rehabilitative, or restitutional. Third, "crimes" are so-defined in relation to coercive, punitive laws; but all coercive, punitive laws are immoral; therefore the "crime-&-punishment" legal justice systems in all states anywhere in the world are also immoral. Let's call this view, *Existential Kantian Cosmopolitan Social Anarchism About Crime-&-Punishment*. In post-"crime-&-punishment" societies,

(i) institutionalized *forgiveness* would replace institutionalized *vengeance*, (ii) *taking deep moral responsibility* for one's choices and actions, and changing one's life for the better, in pursuit of principled authenticity, would replace institutionalized *guilt*, and (iii) *temporary restraint, for the purposes of last-resort defense, protection, and prevention of harm to people* would replace *prisons*. For more details, see Hanna, *Kant, Agnosticism, and Anarchism*, part 3.

[249] See note 162 above.

[250] See, for example, W. Benjamin, "The Work of Art in the Age of Mechanical Reproduction," available online at URL = <http://web.mit.edu/allanmc/www/benjamin.pdf>; A. Danto, *The Transfiguration of the Commonplace* (Cambridge, MA: Harvard Univ. Press, 1981); R. Otto, *The Idea of the Holy*, trans. J.W. Harvey (2nd edn.; Oxford: Oxford Univ. Press, 1958); and M. Eliade, *The Sacred and the Profane*, trans. W.R. Trask (New York: Harcourt Brace Jovanovich, 1987).

[251] See Hanna, *Kant, Agnosticism, and Anarchism*, section 3.14.

[252] Ibid.

[253] See Kant (*LE* 27: 458-460, 709-710) and (*MM* 6: 442-444). For three different contemporary Kantian ethical approaches to the treatment of non-human non-persons, see A. Wood and O. O'Neill, "Kant on Duties Regarding Nonrational Nature," *The Aristotelian Society, Supplementary Volume LXXII* (Oxford: The Aristotelian Society, 1998), pp. 188-210, and 211-228; and C. Korsgaard, "Fellow Creatures: Kantian Ethics and Our Duties to Animals," *Tanner Lectures 2004*, available online at URL = <http://www.tannerlectures.utah.edu/lectures/documents/volume25/korsgaard_2005.pdf>

[254] See, for example, Regan, *The Case for Animal Rights*, pp. 174-185.

[255] B. Hooker, "Sacrificing for the Good of Strangers—Repeatedly," *Philosophy and Phenomenological Research* 59 (1999): 177-181, at 177.

[256] Singer, "The Singer Solution to World Poverty," p. 144.

[257] On the important distinction between (i) *deep* and (ii) *shallow* responsibility, both moral and non-moral, see Hanna, *Deep Freedom and Real Persons*, esp. chs. 4-5.

[258] See note 298 below for the rationale behind this particular element of the principle.

[259] Ibid.

[260] See R. Hanna, "Morality *De Re*: Reflections on the Trolley Problem," in J.M. Fischer and M. Ravizza (eds.), *Ethics: Problems and Principles* (New York: Harcourt, Brace, and Jovanovich, 1991), pp. 318-336; and R. Hanna, "Participants and Bystanders," *Journal of Social Philosophy* 24 (1993): 161-169.

[261] See T. Nagel, "Moral Luck," in Nagel, *Mortal Questions*, pp. 24-38; B. Williams, "Moral Luck," in B. Williams, *Moral Luck* (Cambridge: Cambridge Univ. Press, 1981), pp. 20-39; D. Domsky, "There Is No Door: Finally Solving the Problem of Moral Luck," *Journal of Philosophy* 101 (2004): 445-464; and D. Statman, "Doors, Keys, and Moral Luck: A Reply to Domsky," *Journal of Philosophy* 102 (2005): 422-436.

[262] For an engaging and poignant treatment of this general issue, see L. MacFarquhar, *Strangers Drowning: Impossible Idealism, Drastic Choices, and the Urge to Help* (New York: Penguin, 2016).

[263] P. Foot, "The Problem of Abortion and the Doctrine of the Double Effect," in P. Foot, *Virtues and Vices* (Berkeley, CA: Univ. of California Press, 1978), pp. 19-32.

[264] There are some substantive issues here concerning the justifiability and viability of using commonsense moral intuitions together with the Rawlsian method of Wide Reflective Equilibrium. Foot, Thomson, Kamm, Boonin, and many other contemporary philosophical ethicists take it for granted. But others—for example, Unger, and many contemporary philosophers working in Experimental Philosophy, primed by recent work in cognitive neuroscience—do not, and indeed reject it. My own view on the nature of intuitions in general and on the nature of moral intuitions in particular is orthogonal to both sides of this debate—see, for example, Hanna, *Cognition, Content, and the A Priori*, ch. 7. At the same time, however, I have some skeptical worries about the method of common sense intuitions + reflective equilibrium, that are not dissimilar to those aired by Experimental Philosophers.

In any case, it must be fully conceded that there are some reasonable people who reject the initial moral-intuitional data of The Trolley Problem, The Self-Defense Problem, and The Famine Relief Problem. For example, as anyone who teaches Introductory Ethics knows, some reasonable people see no relevant moral differences whatsoever between *Bystander at the Switch* and *Fat Man*, or between *Pond* and *Envelope*, and also see no relevant moral identities between *Villainous Aggressor* and *Innocent Aggressor*, or between *Villainous Aggressor* and *Innocent Threat*. I also think that even despite their being reasonable people, they are nevertheless *mistaken* here, and in the grip of a cognitive illusion. But that is a long story for another day.

[265] See also P. Foot, "Killing and Letting Die," in S. Cahn and P. Markie (eds.), *Ethics: History, Theory, and Contemporary Issues* (3rd edn.; New York: Oxford Univ. Press, 2006), pp. 783-788.

[266] See J.J. Thomson, "Killing, Letting Die, and the Trolley Problem," in J.M. Fischer and M. Ravizza (eds.) *Ethics: Problems and Principles* (New York: Harcourt, Brace, and Jovanovich, 1991), pp. 67-77; and J.J. Thomson, "The Trolley Problem," in Fischer and M. Ravizza (eds.) *Ethics: Problems and Principles*, pp. 279-292.

[267] Thomson, "The Trolley Problem," p. 284.

[268] Thomson, "The Trolley Problem," p. 285.

[269] Thomson, "The Trolley Problem," p. 288.

[270] For an earlier version of this case—approached in three steps—see Hanna, "Morality *De Re*: Reflections on the Trolley Problem," pp. 325-326. When I wrote this paper (1988 or 1989), I did not see that this case is also a clear counterexample to the unrestrictedly universal, orthodox Kantian moral principle which says that it is impermissible to treat someone as a mere means. But in the meantime, that point was made very effectively by Kamm and Parfit: see note 271 below. Still, neither Kamm nor Parfit notices the radical difference that is made

by adding an "other things being equal"/*ceteris paribus* clause to first-order Kantian moral principles. In fact, Kamm/Parfit counterexamples obtain only when other things *aren't* equal, hence they're *not* counterexamples to the first-order *ceteris paribus* version of the "mere means" principle.

[271] See also Kamm, *Intricate Ethics*, ch. 5; and Parfit, *On What Matters*, vol. 1, ch. 9.

[272] Rawls's veil of ignorance is of course simply a sub-routine within a larger rational deliberative mechanism for choosing moral principles. But to the extent that any higher-level or Kantian real human person is asked to provide actual or possible rational consent to some treatment, she puts herself behind a self-imposed veil and morally considers her own situation *as if* it were that of any other member of The Realm of Ends.

[273] According to the de vs de dicto distinction, a proposition is *de re* if it is *referentially committed to the existence of a certain unique actual object* that also, as it happens, bears a certain descriptive profile, but *de dicto* if that *proposition refers to any object whatsoever that bears that descriptive profile*, whether or not such an object actually exists.

[274] See Kamm, *Intricate Ethics*, pp. 24-25, and chs. 3-5.

[275] Kamm, *Intricate Ethics*, ch 1., esp. at pp. 11-14.

[276] See section 3.2 above; and also Hanna, *Deep Freedom and Real Persons*.

[277] Kamm, *Intricate Ethics*, p. 5 and ch. 14.

[278] See, for example, J. Dower, *War Without Mercy: Race and Power in the Pacific War* (New York: Pantheon Books, 1986).

[279] See, for example, *Human Rights Watch*, available online at URL = <http://www.hrw.org/>.

[280] See, for example, J. Perry, "The Problem of the Essential Indexical," *Noûs* 13 (1979): 3-21; and R. Hanna, "Direct Reference, Direct Perception, and the Cognitive Theory of Demonstratives," *Pacific Philosophical Quarterly* 74 (1993): 96-117.

[281] See, for example, Hanna, *Cognition, Content, and the A Priori*, ch. 2.

[282] See, for example, D. Kaplan, "Demonstratives: An Essay on the Logic, Metaphysics, Semantics, and Epistemology of Demonstratives and Other Indexicals" and "Afterthoughts," both in J. Almog et al. (eds.), *Themes from Kaplan* (New York: Oxford Univ. Press, 1989), pp. 481-563 and 565-614.

[283] See Hanna and Maiese, *Embodied Minds in Action*, section 5.3; Maiese, *Emotions, Embodiment, and Cognition*, esp. ch. 5.

[284] Thomson, "Self-Defense," p. 283.

[285] By "personhood-destroying suffering" I mean suffering that is so intense, so prolonged, and so unrelievable that only death will prevent the real person's permanently losing her rational capacities altogether as a result of this suffering. See also section 6.9 below.

[286] Thomson, "Self-Defense," p. 284.

[287] See also Thomson, "Self-Defense," p. 287. There is a subtle but important refinement here that I have marked by calling my case *Innocent Attacker*. As I noted above, Thomson actually calls her third case *Innocent Threat*. But what is philosophically odd about her case is that she stipulates that a villain has *pushed* the Fat Man. So by my understanding of her *Innocent Aggressor* cases, since the Fat Man is merely an innocent tool of the villainous aggressor, her *Innocent Threat* case is actually an *Innocent Aggressor* case. Nevertheless, reading Thomson charitably, I am convinced that she intends the third case as a villain-less *Innocent Attacker* case, since the villain plays no role in her discussion once she has described the case. Many thanks to Adam Betz for suggesting the term "innocent attacker" to me.

[288] Many thanks again to Adam Betz, this time for getting me to think about this illuminating variant.

[289] There is a further variant on *Armed Innocent Defensive Attacker* in which the Fat Man has a gun but I am not myself a Fat Man. So by shooting me in this scenario, he will not save his own life, but instead only kill me in order to prevent his being blown up, as opposed to his dying from the impact when he hits the ground, or perhaps just because, seeing that he is going to die no matter what, he simply wants to "take me with him." Since in this variant the Fat Man cannot kill me in self-defense but rather only kill me as an offensive act, I will call this variant *Armed Innocent Offensive Attacker*. This seems obviously morally impermissible, precisely because the Fat Man is violently stopping me from morally permissibly saving my own life. So he is treating me without my actual or possible rational consent, which I would surely refuse, and thus he is harming me by violating my dignity.

[290] See Hanna, *Kant, Agnosticism, and Anarchism*, section 3.8.

[291] Ibid.

[292] See note 181 above; and also Singer, "The Singer Solution to World Poverty."

[293] P. Unger, *Living High and Letting Die: Our Illusion of Innocence* (New York: Oxford Univ. Press, 1996). See also R. Hanna, "Must We Be Good Samaritans? On Unger's *Living High and Letting Die*," *Canadian Journal of Philosophy* 28 (1998): 453-470.

[294] Unger, *Living High and Letting Die*, pp. 24-25.

[295] This is to accommodate the fact that there may be several people who are equally close to the endangered real person, but either not all of them, or none of them, are able to save the person; whereas some person who is slightly further away, but also still quite close, is (also) able to save the endangered person. The intention of the principle is to require joint satisfaction, by a single person, of all three factors I specified, in order to be obligated;

and also in cases in which a person is not obligated, to require joint non-satisfaction by the same person. Obviously it is *permissible* for anyone, whether close by or far away, to try to save the endangered person. Thanks to Kelly Vincent for urging me to be more explicit and precise about these points.

[296] Singer, "The Singer Solution to World Poverty," p. 144.

[297] See note 261 above.

[298] See Thomson, "A Defense of Abortion," at pp. 11 and 17. The italicized phrase is of course Kierkegaard's; and Thomson's original case involved Ms. Fonda's father, Henry Fonda.

[299] See MacFarquhar, *Strangers Drowning*, p. 101.

[300] See also S. Buss, "Needs (Someone Else's), Projects (My Own), and Reasons," *Journal of Philosophy* 103 (2006): 373-402.

[301] Epicurus, *Letter to Menoeceus*. Available online at URL = <http://www.epicurus.net/en/menoeceus.html>.

[302] Lucretius, *De Rerum Natura*, as quoted in S. Luper, "Death," section 3.2, *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), E.N. Zalta (ed.), available online at URL = <http://plato.stanford.edu/archives/win2014/entries/death/>.

[303] F. Nietzsche, *The Portable Nietzsche*, trans. W. Kaufmann (Harmondsworth, Middlesex: Penguin, 1983), pp. 101-102 (*Gay Science*, # 341), italics in the original.

[304] Wittgenstein, *Tractatus Logico-Philosophicus*, p. 185.

[305] D. Thomas, "Do Not Go Gentle into that Good Night," in O. Williams (ed.), *The Pocket Book of Modern Verse* (New York: Washington Square, 1973), p. 486.

[306] See note 248 above; and Hanna, *Kant, Agnosticism, and Anarchism*, section 3.10.

[307] See for example, Hanna, *Deep Freedom and Real Persons*, ch. 2.

[308] For a sharply contrasting approach, see Parfit, *On What Matters*, vol. 2, ch. 36. Basically, the sharp contrast here is between my Kant-inspired *Kierkegaardianism* on the one hand, and Parfit's Kant-inspired *Sidgwickianism* on the other.

[309] See Hanna, *Deep Freedom and Real Persons*, chs. 6-7.

[310] D. Suits, "Why Death is Not Bad for the One Who Died," *American Philosophical Quarterly* 38 (2001): 69-84.

[311] See Hanna and Maiese, *Embodied Minds in Action*, pp. 20 and 207.

[312] Nevertheless, I also think that, in principle, there is no good reason *whatsoever* why contemporary philosophers should not be helping to create the philosophy of the future by using new, experimental presentational formats in order to explore these *us*-centered themes, and many others, more deeply. See R. Hanna, "Life-Changing Metaphysics: Rational Anthropology and its Kantian Methodology," in G. D'Oro and S. Overgaard (eds.), *The Cambridge Companion to Philosophical Methodology*, (Cambridge: Cambridge Univ. Press, 2016), pp. 201-226, section 4; and Hanna, "Preface and General Introduction to THE RATIONAL HUMAN CONDITION," THE RATIONAL HUMAN CONDITION, Vol. 1, section 1.4.

[313] W. Shakespeare, *Hamlet* (1607), act V, scene ii.

[314] See Hanna, *Deep Freedom and Real Persons*, chs. 6 and 7.

[315] B. Williams, ""The Makropulos Case: Reflections on the Tedium of Immortality," in B. Williams, *Problems of the Self* (Cambridge: Cambridge Univ. Press, 1973), pp. 82-100.

[316] J.M. Fischer, "Why Immortality is Not So Bad," *International Journal of Philosophical Studies* 2 (1994): 257-270.

[317] See Hanna, *Deep Freedom and Real Persons*, chs. 6-7; and chapter 3 above.

[318] S. Critchley, *Infinitely Demanding* (London: Verso, 2007), pp. 3-6.

[319] T. Nagel, "Death," in Nagel, *Mortal Questions*, pp. 1-10.

[320] See note 310 above.

[321] So, other things being equal, it is *not* better never to have existed. For an opposing view, see D. Benatar, *Better Never to Have Been: The Harm of Coming into Existence* (Oxford: Oxford Univ. Press, 2008).

[322] Suits, "Why Death is Not Bad for the One Who Died," p. 79.

[323] Shakespeare, *Hamlet*, act III, scene i.

[324] See Hanna, *Deep Freedom and Real Persons*, section 3.3.

[325] See, for example, R. Solnit, *A Paradise Built in Hell: The Extraordinary Communities That Arise in Disaster* (London: Penguin, 2009); and MacFarquhar, *Strangers Drowning*.

[326] See Adams, *The Hitchhiker's Guide to the Galaxy*. The digit '42' is the justly famous answer provided by the supercomputer Deep Thought, after 7.5 million years of computing, to the Ultimate Question of the Meaning of Life, the Universe, and Everything. Sadly, however, the Ultimate Question itself remains unknown.

[327] (Dir. C. Eastwood, 2004).

[328] J. Rachels, "Active and Passive Euthanasia," in Cahn and Markie (eds.), *Ethics: History, Theory, and Contemporary Issues*, pp. 779-783.

[329] This seems particularly likely in view of the fact that, other things being equal, higher-level or Kantian real human persons ought to regard their future rational selves as cognitive experts with respect to present facts. See S. Evnine, *Epistemic Dimensions of Personhood* (Oxford: Oxford Univ. Press, 2008), ch. 5.

[330] P. Foot, "Euthanasia," in Foot, *Virtues and Vices*, pp. 33-61.

[331] Sometimes, of course, children or adolescents commit suicide: in almost all such cases, it is simply tragic, not

immoral.

[332] See note 88 above.

[333] Ibid.

[334] Directed by Billy Wilder (1944).

[335] See note 81 above.

[336] Aristotle, *Nicomachean Ethics*, trans. T. Irwin (Indianapolis, IN: Hackett, 1985), 1100a10-32 and 1101a11-1101b9 ; and Nagel, "Death," pp. 4-7.

[337] See note 302 above; Nagel, "Death," p. 7; and Kamm, *Morality, Mortality*, vol. 1, chs. 2-3.

[338] See note 305 above.

[339] R.M. Rilke "Archaic Torso of Apollo," in *The Selected Poetry of Rainer Maria Rilke*, trans. S. Mitchell (New York: Vintage, 1989), p. 61, lines 13-14.

INDEX