

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600**

UMI[®]

Many-valued Morality

by

Merav Brodetz

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The City University of New York

2000

UMI Number: 9986306

UMI[®]

UMI Microform 9986306

Copyright 2000 by Bell & Howell Information and Learning Company.

All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

Bell & Howell Information and Learning Company

300 North Zeeb Road

P.O. Box 1346

Ann Arbor, MI 48106-1346

Abstract**Many-valued Morality**

by

Merav Brodetz

Adviser: Professor Michael Levin

Punishments and rewards are proportionate to actions' *degree* of wrongness and rightness, which means that gradation is a datum about morality. In my dissertation I argue that this datum cannot be captured by existing—that is, two- and three-valued—moralities, but *can* be accommodated by a many-valued system. I then develop a many-valued morality and obtain several important results. Among the latter are (i) a pluralist (multi-standard) ethics: by contrast to both two- and three-valued moralities, a many-valued framework allows one to always weigh more than a single standard when evaluating actions; (ii) a sympathetic reading of Kant's 'lack of moral worth': from a many-valued perspective, this notion (which many deem problematic) is but a third moral value (in addition to 'right' and 'wrong'), and reflects an attempt to capture moral gradation; and (iii) a new outlook on an aspect of the dispute between utilitarianism and deontological justice: whether killing a few to save many is unjust is a matter of degree: whether utilitarianism conflicts with justice over this issue is *not*.

In memory of my grandparents. Moshe and Bronya Marmurstein and Arieh and Bracha Rosenthal. and in memory of my father in law. Baruch Sarel.

Acknowledgments

For their helpful comments I am grateful to Ofer Fein, Inbal Heth, David Heyd, Tziporah Kasachkoff, Asa Kasher, Douglas Lackey, Yoni Lebowitsch, Edo Likhovsky, Ha'im Marantz, Assaf Peretz, Graham Priest, Peter Simpson, Achille Varzi, my students at Baruch College and John Jay College and audiences at Rutgers University, Ben-Gurion University and the 28th Conference on Value Inquiry.

Eli Dresner discussed with me every major aspect of the project. His insightful suggestions have left their mark on significant portions of the end result.

Jonathan Adler, Julia Driver, Rohit Parikh and Steven Ross commented on drafts of this work and held numerous conversations and e-mail correspondences with me. Their perceptive criticisms and suggestions have greatly influenced my writing, their care and concern—my endurance and well being.

Michael Levin has been a *model* adviser, embodying every trait a student could wish for. He helped me steer away from an unpromising topic and focus on this unusual one instead. Then, over a period of two and a half years, he showered on the dissertation an enormous number (in the thousands!) of invaluable comments—all of which he e-mailed to me within *days* of receiving the drafts—without ever threatening my sense of intellectual liberty. My debt to Michael is much greater than any few lines could convey.

Through their persistent support, the members of my committee and Tziporah Kasachkoff have been able to reverse the traumatic impact of figures whose names belong under the title 'disacknowledgments.'

My parents, Gadi and Hannah, have been remarkably generous and courageous, especially considering the fact that they would have *much* preferred investing their limited resources in a job-securing course of studies.

Running the risk of availing myself to the mother of all clichés. I must say that words—especially those of a foreign tongue—will not capture the love and gratitude I feel towards Dan Sarel, my partner and friend of over nine years. He made major contributions to all aspects of the dissertation. And he has always been so good to me that the very thought of it fills my eyes with tears.

Preface

In his *Groundwork of the Metaphysic of Morals* Kant makes such claims as the following. Helping others when one finds “inner pleasure in spreading happiness...and can take delight in the contentment of others”¹ certainly accords with duty (it wouldn’t occur to us to forbid such conduct), yet the action lacks genuine moral worth. On the other hand, helping others when one’s mind “is overclouded by sorrows of his own which extinguished all sympathy with the fate of others”² has true moral worth. A great many sensitive readers of Kant believe that this position of his is troubling, as it implies that resentfully performing a dutiful action (saving the life of a person one loathes) has true moral worth, whereas lovingly or joyfully performing a similar action (saving the life of a loved one) altogether lacks moral worth.

I accordingly used to designate the position ‘problematic’ when presenting it to students. Not for long though. At one point, teaching the *Groundwork* coincided with my taking graduate courses on vagueness and modal logic. It then occurred to me that ‘lack of moral worth’ plays in Kant’s system roughly the role that ‘neither true nor false’ plays in many-valued logics. In introducing ‘lack of moral worth,’ Kant in fact gestures towards a moral theory capable of capturing the intuition that right acts don’t all enjoy a uniform degree of rightness: some are *more right* than others (due to involving sacrifice on the agent’s part). Since the intuition in question seemed undeniable (rightness no doubt admits of degrees), I concluded that Kant’s position is

¹ 66: 398. (Page numbers refer to H. J. Paton’s translation and the Prussian Academy of the Sciences edition, respectively.)

² *Ibid.*, *ibid.*.

anything but problematic.

From here to the point of recognizing that there's a need for a continuum-valued morality, the road was short. At first I did little to develop this notion, as I was preoccupied with writing a dissertation prospectus on causation under Professor Michael Levin's supervision. Soon after however, I began laboring on a paper entitled "Many-valued Moral Systems" for a conference on ethical narratives. Throughout this period of time, I exchanged ideas on both causation and many-valued morality with Eli Dresner.

In the meanwhile, the prospectus on causation didn't quite take off. I complained about my *tzures* to Eli. Quoting one of his professors at the time, he said that a dissertation requires but a single fruitful idea, "which you've got: many-valued morality." I declined. A few weeks later, as the deadline for submitting "Many-valued Moral Systems" was nearing, I sent the paper to Professor Levin for his comments. Right at the start of our next meeting, Professor Levin—holding the paper in his hand—asked: "Why don't you write your dissertation on *this*?" By the evening I converted from causation to many-valued morality. I have never looked back since.

Table of Contents

Illustrations	x
Chapter 1: Introduction and Preliminaries	1
Part I: Many-valued Morality	17
Chapter 2: ‘Not’	19
Chapter 3: ‘Or’	60
Chapter 4: ‘And’	86
Chapter 5: ‘By’	119
Chapter 6: Concluding Remarks	143
Part II: Applications	152
Chapter 7: Pluralism or Multi-standard Ethics	153
Chapter 8: Kant’s ‘Lack of Moral Worth’	165
Chapter 9: Utilitarianism and Deontological Justice	184
Bibliography	209

Illustrations

1.1. 'Not': Right Actions	25
1.2. 'Not': Wrong Actions	36
1.3. 'Not': Graphic Representation	38
5.1. 'By'	119
8.1. Kant's Three-valued Morality	169

Chapter 1: Introduction and Preliminaries

1. The Data

Murderers are sentenced to life imprisonment or even to execution, but shoplifters are merely fined. People who in addition to their regular jobs engage in volunteer work are highly praised, yet keeping a trivial promise receives no special mention. In short, rewards and punishments are proportionate to the degree of rightness and wrongness of actions.

This practice indicates that rightness and wrongness—just as are sweetness or wealth—are properties that admit of degrees (scalar). A second—closely related—indication for moral gradation: in various legal systems, crimes are classified according to their gravity. Many US jurisdictions for instance distinguish between felonies and misdemeanors. Felonies are the more serious crimes: they are typically punishable by a term of imprisonment of not less than one year. Misdemeanors on the other hand are often defined as punishable only by fines or by short terms of imprisonment in local jails.

Thirdly, in various countries, mediation is increasingly replacing trials as means for conflict resolution. A trial is an essentially binary proceeding: the plaintiff wins and the defendant loses or vice versa. It therefore entails that one's conduct is either entirely right or entirely wrong. Mediation by contrast strives to culminate in a compromise. Thus it allows that each side is right to a degree.

Fourth, we say such things as that certain acts are *more or less* right. Or that certain acts—though certainly right—could have been *more right* had they been carried out differently.

A fifth indication concerns conflicts of duties. Imagine that in order to moderately prolong a life, a doctor must lie. The duty to prolong lives thus conflicts with the duty of truth-telling. A rational resolution to the conflict inevitably involves determining which of the two is the more pressing duty. Or which is likely to produce a *higher degree of rightness*.

Finally there are views on afterlife.¹ Christian theologians in the Middle Ages held that some souls—for example, souls of dead unbaptized infants—dwell in between heaven and hell (*in limbo*), because they deserve neither heaven (their original sin has not been washed away by baptism), nor hell (they have died without actual sin). Thus, the theologians saw a need for a third moral value—‘neutral’—in addition to ‘right’ and ‘wrong.’ (‘Purgatory’ is of course also a notion that indicates recognition of moral gradation on the part of theologians.)

2. Moral Gradation Explained

As to *why* there is moral gradation, the explanation is I think quite simple. In evaluating actions, we tend to employ essentially two standards, and they both construe rightness and wrongness in terms of scalar properties. The first focuses on what is brought about by actions (consequences or effects). It states roughly that an act is right or wrong if it promotes pleasure or produces pain, respectively. The consequentialist standard then construes rightness and wrongness in terms of pleasure and pain, and these—clearly—admit of degrees.

The second standard centers on what brings actions about (motives or causes). It says roughly that an action is right or wrong depending on whether the agent has good or

¹ My use of ‘finally’ notwithstanding, I don’t wish to imply that the list I’ve compiled in this Section is exhaustive.

bad motives. A natural way to cash out this claim can be found in the legal system. There, a person is judged to have *mens rea* (i.e., “guilty mind” or bad intentions) only when he has acted voluntarily and had foresight of the consequences. The motive-oriented standard then construes rightness and wrongness in terms of foresight and voluntariness, both of which are scalar.

As to foresight, since it’s a function of—among other things—cognitive development, which admits of degrees, it too admits of degrees. Thus, a two year old, a seven year old, and a (mentally healthy) twenty year old are—respectively—incapable, partly capable, and fully capable of having foresight of the consequences of their actions. (Accordingly, if they are to set a house on fire, they would—respectively again—not be held responsible, be held partly responsible, and be held fully responsible for their action.) As for voluntariness, the conditions which undermine it—namely, coercion, compulsion, duress, posthypnotic suggestion, and so on—admit of degrees. Thus, a person whose life is threatened is more coerced to perform an act than a person whose job is threatened.

3. Two-valued Moralities: Absolutism

There is then a sound basis for holding that gradation is a datum about morality. Since theories should incorporate the data about their subject matter, our moral theory ought to capture moral gradation.

Yet moral gradation cannot be captured (at least not *fully*) by the theories found in the philosophical literature: that is, two- and three-valued moralities. A two-valued framework—such as absolutism or utilitarianism—allows one to assign to actions just one of only two moral values: ‘right’ and ‘wrong.’ Such a system cannot accommodate

moral gradation *at all*. For while allowing one to say for instance that murder and theft are both wrong, it doesn't allow saying that murder is *more wrong* than theft.

I said that utilitarianism and absolutism each constitute a two-valued morality. I should now like to elaborate on and qualify this claim. Absolutism is discussed in this Section, utilitarianism in the next.

Consider first Kant's absolutism. According to Kant, if willing an action is consistent with willing that everyone in the same situation would do the same, the action is permissible. If willing an action is *inconsistent* with willing that everyone under equivalent circumstances would do the same—that is, if contradictory wills result—the action is forbidden. I take it that an act is obligatory if willing its omission is incompatible with willing that everyone under equivalent conditions would also omit the act.

Kant holds then that the deontic status of acts rests on logical contradiction. Since logical contradiction is an all-or-nothing (binary) phenomenon (there are no *degrees* of contradiction and consistency),² Kant's absolutism entails that obligatoriness too is a binary property. Now if a property *P* is binary, the objects to which it applies are all equally *P*. For example, all odd numbers are equally odd: the number 9 doesn't enjoy a higher degree of oddness than the number 5 and vice versa. Therefore, Kant's absolutism entails that all duties are equally obligatory. Or that obligatoriness doesn't admit of degrees.

As moral dilemmas effectively demonstrate however, obligatoriness *does* admit of degrees. If *A* and *B* are conflicting duties, and the agent recognizes that (say) refraining

² This is not deny a point that Ruth Barcan Marcus makes in her "Moral Dilemmas and Consistency": namely, that *systems* (collections of claims) may have varying degrees of *potential* to entail contradictions.

from *A* would have much graver consequences than refraining from *B*. then—intuitively—*A* is *more obligatory* than *B*. Alternatively, if obligatoriness weren't scalar—if all duties were equally obligatory—moral dilemmas would never require deliberation (it would make no moral difference which horn agents opt for). Intuitively however, moral dilemmas do require deliberation, if not always then at least frequently.

Obligatoriness then admits of degrees. Moreover, as is evident from the foregoing discussions, both obligatoriness and rightness are scalar properties because goodness and badness are. This means that obligatoriness and rightness are correlated. Or that obligatoriness is a scalar property iff rightness is. Thus, if obligatoriness doesn't admit of degrees, rightness doesn't either. Since according to Kant's absolutism obligatoriness doesn't admit of degrees, the theory implies that rightness and wrongness too don't admit of degrees.³

Turning now to religious absolutism, this theory states that actions are right iff they accord with the divine command. Since being in accordance with the divine command is an all-or-nothing property (actions either accord with the divine command or they don't), religious absolutism entails that rightness and wrongness rest on a binary property. Like Kant's absolutism therefore, religious absolutism has no room in it for moral gradation.

As to why absolutism excludes the notion that rightness and wrongness admit of degrees, I address this issue in Appendix to Chapter 2, Section 4.

³ Further discussion of this matter appears in Chapter 8, Section 12.

4. Two-valued Moralities: Utilitarianism

Utilitarianism requires agents to always produce the best (some say: optimal⁴) overall outcome. It therefore entails that an act is right iff it maximizes (or optimizes) the general good; failing to maximize (optimize) the general good is wrong. Since each act either maximizes (optimizes) the general good or not (maxima and optima don't admit of degrees), maximization (optimization) versions of utilitarianism imply that each and every act is either right or wrong: there are no *degrees* of rightness and wrongness.⁵ Thus, maximization (optimization) versions of utilitarianism constitute two-valued moralities. Henceforth I shall refer to any such theory as two-valued utilitarianism.

Four brief comments. First of all, the theory *necessarily* implies that rightness doesn't admit of degrees; yet it implies that wrongness doesn't admit of degrees only *contingently* (or, if you will, conversationally or contextually). Two-valued utilitarianism explicitly states that acts are right iff they maximize (optimize) utility: since maxima and optima don't admit of degrees, rightness doesn't either. On the other hand, the theory doesn't state that (say) acts are wrong iff they *minimize* utility (if it did, it *would* necessarily entail that wrongness doesn't admit of degrees, for minima don't admit of degrees). It says only that acts are wrong iff they *fail* to maximize utility. Thus, nothing prevents utilitarians from holding for example that (1) in case an act produces utility, its wrongness is inversely related to the degree (in percentages) to which it fails to maximize (optimize) utility, and (2) in case an act generates disutility, its wrongness is directly related to the degree of disutility it generates.

⁴ See Charles Fried's *An Anatomy of Values*, 170-6 and Amartya Sen's "Utilitarianism and Welfarism," 470-1.

⁵ Its drawbacks notwithstanding, two-valued utilitarianism does answer an important theoretical need. More on this matter in Chapter 3, Section 5.

While being incompatible with scalar rightness then, two-valued utilitarianism is not incompatible with scalar wrongness. To the best of my knowledge however, utilitarians have manifested no interest in supplementing their theory with a definition of scalar wrongness, thereby (conversationally or contextually) implying that—like rightness—wrongness too is a binary property.

Secondly, supposing that according to two-valued utilitarianism *both* rightness and wrongness are binary properties, the theory entails not only that as long as each maximizes the general good, there exists no moral difference between saving a life and keeping a trivial promise. Or that as long as each fails to maximize overall goodness, murder and theft are equally wrong. It also entails that as long as each fails to produce the best overall outcome, keeping a trivial promise and murder are equally wrong.

Thirdly, both Bentham and Mill oscillate between two- and many-valued utilitarianism. Mill calls the principle of utility ‘the *greatest* happiness principle’⁶ thereby suggesting that he is committed to two-valued utilitarianism. On the other hand he writes that

...actions are right *in proportion* as they tend to promote happiness, wrong as they tend to produce the reverse of happiness.⁷

Since happiness may be promoted or opposed to various degrees, this passage in fact encapsulates a many-valued version of utilitarianism.

Bentham maintains that the principle of utility

states the *greatest* happiness of all those whose interest is in question, as being...the only right and proper...end of human action...⁸

⁶ *Utilitarianism*, 137.

⁷ *Ibid.*, *ibid.*. Italics mine.

⁸ *Introduction to the Principles of Morals and Legislation*, Ch. 1, Section 1, fn.. Italics mine.

At the same time he argues that

By the principle of utility is meant that principle which approves or disapproves of every action whatsoever, according to the tendency which it appears to have...to promote or oppose...happiness.⁹

Now I maintained that existing theories cannot accommodate moral gradation.

Given many-valued utilitarianism, this statement is somewhat misleading. ‘Somewhat.’ because in the first place, many-valued utilitarianism is a theoretical non-starter, as it implies that producing a crumb of goodness (for instance, preventing a minor harm) is right even if a *much* better alternative (saving a life for example) is available to the agent. Secondly, many-valued utilitarianism is under-described to such an extent that its very existence has been generally overlooked. For all intents and purposes therefore, the class of existing moralities includes only two- and three-valued frameworks.¹⁰

The fourth and final remark concerns Michael Slote’s satisficing utilitarianism. (Two-valued) utilitarianism requires agents to *always* maximize utility. It therefore entails that (say) tending to one’s garden or stamp collection is wrong. It entails in other words that agents should abandon their deepest commitments and projects whenever these fail to maximize the general good. Which is implausible. To shield consequentialism against this and related concerns, Slote allows that in order for their acts to be morally right, agents need not always aim at producing the best possible

⁹ *Ibid.*, Ch. 1, Section 2.

¹⁰ The first and third comments indicate then that while certainly an instance of two-valued morality, two-valued utilitarianism isn’t as good an instance of it as absolutism. This should come as no surprise. Utilitarianism after all revolves around scalar notions such as pleasure and pain.

outcome: their acts might qualify as morally right through having *good enough*—through satisficing—consequences.¹¹

Slote suggests that a good enough consequence amounts to some sort of percentage of the best results attainable by the agent.¹² This means that an act is right if it produces at least the required amount of goodness, wrong otherwise. Or that an act is right if it crosses a certain threshold, wrong otherwise. No less than traditional versions of utilitarianism then, satisficing utilitarianism is a two-valued moral framework incapable of capturing moral gradation.

5. Three-valued Moralities

I argued that a two-valued framework allows one to assign to actions just one of only two moral values: ‘right’ and ‘wrong’; such a system cannot accommodate moral gradation *at all*.

A three-valued morality (such as one of the theories advanced in Kant’s *Groundwork*; see Chapter 8) allows one to assign to actions any of three moral values: ‘right,’ ‘wrong,’ and ‘neutral.’ Such a framework can capture moral gradation *to some extent*, as it allows one to say for example that the “conduct” of dead unbaptized infants was neither right nor wrong (neutral). But a three-valued system cannot *fully* represent moral gradation, because it doesn’t allow saying things such as that although murder and shop lifting are both wrong, the latter is much *more wrong* than the former.

The difficulty afflicting two- and three-valued moralities can be cast in terms of information loss. Take two-valued moralities. In such systems, all right actions are indiscriminately categorized as right, all wrong actions as wrong. We thus lose

¹¹ *Common-sense Morality and Consequentialism*, Chapter III.

¹² *Ibid.*, 52.

information about the moral difference between saving a life and keeping a trivial promise. Or between murder and theft. The same applies (*mutatis mutandis*) to three-valued theories.

6. Many-valued Moralities

By contrast to both two- and three-valued moralities, a many-valued system can *fully* capture moral gradation. It allows saying that murder is *more wrong* than theft, that keeping a trivial promise is not *as right as* saving a life, and so on.

In addition, many-valued morality has several important ramifications for major issues in ethics. I present three of them in Part II. They include (i) a pluralist (multi-standard) ethics: by contrast to both two- and three-valued moralities, a many-valued framework allows one to always weigh more than a single standard when evaluating actions; (ii) a sympathetic reading of Kant's 'lack of moral worth': from a many-valued perspective, this notion (which many deem problematic) is but a third moral value (in addition to 'right' and 'wrong'), and reflects an attempt to capture moral gradation; and (iii) a new outlook on an aspect of the dispute between utilitarianism and deontological justice: whether killing a few to save many is unjust is a matter of degree: whether utilitarianism conflicts with justice over this issue is *not*.

Assessing many-valued morality according to both the considerations that motivate it and its consequences for ethics naturally leads to the conclusion that such a system should be constructed. I take on this project in Part I. It joins an ongoing (though regrettably rather marginalized) philosophical effort of utmost significance: namely, introducing precision—which is essential for consistency, which in turn is essential for justice—into processes of decision making in ethics.

7. Preliminaries: Central Assumptions

Having introduced the rationale behind—and nature of—this work, I now turn to preliminary comments. This Section concerns methodology and central assumptions. In Section 8 I distinguish rightness from goodness. The extent to which the present project is descriptive is the subject matter of Section 9.

Morality and Law As indicated by previous discussions,¹³ I often take my cue concerning moral intuitions from well-established legal practices. Two assumptions motivate this methodological choice. The first is that rather than being accidentally related to morality, the legal system necessarily *reflects* moral intuitions. (That is a crude way of formulating the credo of contemporary “natural law” theories. The view that law and morality are related only loosely and accidentally was espoused by traditional legal positivists, such as John Austin.)

Secondly, well-established legal practices may be considered just, as they have survived a process of “natural selection” or ongoing evaluation. (This claim echoes of course the thought encapsulated in John Rawls’s notion of reflective equilibrium: our general moral principles may be regarded as justified if they survive a prolonged process of examination and re-examination against the intuitive moral judgments which we make in particular cases.¹⁴) The moral intuitions derived from well-established legal practices are therefore unlikely to be seriously confused or misguided.

Objective Consequences For the most part I shall assume that agents are responsible only for foreseen (or subjective) consequences of their acts: actual (objective) consequences are morally *irrelevant*. I shall assume in other words that rightness is a

¹³ Sections 1-2.

¹⁴ *A Theory of Justice*, 48-51.

function of *expected* good consequences, wrongness—of *expected* bad consequences: whether a well-intended act in fact results in good consequences is irrelevant: if—in attempt to prevent injuries among them—an industrialist provides her employees with protective masks that end up harming them, her act is right: ditto with respect to ill-intended acts and bad outcomes: attempted murder is as wrong as murder.

This assumption appears to fly in the face of my methodology. Strict liability statutes—which state that one is responsible for *unforeseen* damage done by things in one's control (such as a car or piece of equipment)—are proliferating throughout the world. (By contrast to strict liability, *fault* liability involves informed choice: it presupposes that an agent could choose between performing an action in a perceptibly dangerous way and performing it in some safer way.) Moreover, the law distinguishes between murder and attempted murder, thus allowing that one is responsible for unforeseen good consequences of his acts.

There is however a reason for thinking that strict liability laws not only don't conflict with—but are in fact anchored in—the notion that persons should be held responsible only for foreseen consequences of their acts. It is that arguably, the knowledge that one is living in a strict liability system might tend to make citizens meticulously careful and thereby reduce the frequency of antisocial acts, might cause them to avoid whole areas of action.¹⁵ That is to say, strict liability systems might ensure that agents take the trouble to foresee—for instance by researching their intended field of activity—consequences that they otherwise wouldn't expect.

However, even if assuming that only foreseen consequences are morally relevant

¹⁵ See Murphy and Coleman, *Philosophy of Law*, 126.

conflicts with common legal practice, I wouldn't be completely breaching my methodological commitment. For in Chapter 7—and in that Chapter alone—I depart from the assumption in question, stipulating instead that both foreseen and actual consequences are morally relevant.¹⁶

Foresight and Voluntariness I argued that rightness and wrongness admit of degrees because they ultimately reduce to scalar properties: namely, pleasure, pain, voluntariness and foresight.¹⁷ With few exceptions however, I shall conduct my discussions assuming that agents act voluntarily and have full foresight of morally relevant consequences of their acts. I shall assume then that rightness and wrongness admit of degrees only because they reduce to pleasure and pain.

Doing and Allowing Powerful arguments to the contrary notwithstanding, I shall assume that the distinction between doing and allowing is morally significant: that (say) killing differs from knowingly letting die. More on this matter in Chapter 2, Section 2 and Chapter 9, Sections 6-7 and Appendix 2.

8. Preliminaries: Rightness and Goodness

If rightness and wrongness reduce to pleasure and pain, one might worry that—since goodness and badness also reduce to pleasure and pain—I in effect identify rightness with goodness and wrongness with badness. (For simplicity's sake, I shall henceforth confine my discussion to rightness and goodness.)

Suppose this is indeed the case. So what? After all, it's not as though there exists already a continuum-valued morality designed to capture the fact that goodness and

¹⁶ I briefly review possible justifications for both assumptions in Chapter 7, Section 7.

¹⁷ Section 2.

badness are scalar properties.¹⁸ I mean: identifying rightness with goodness would in no shape or form render my project superfluous.

Michael Slote's scalar morality merits a comment. Slote maintains, you will recall, that an act is right if it produces sufficiently good consequences. He then points out that there is no natural or non-arbitrary resting place for a distinction between sufficiently- and insufficiently good consequences (would 90% of the best results attainable by the agent be good enough? If so, how could 89% of the best results fail to be good enough?). Consequently, no dividing line between good and bad action would be thought of as corresponding to anything objectively valid.

Slote argues that in light of this observation, talk of good and bad (or right and wrong) actions makes little sense. We should abandon it and instead make only comparative judgments: namely, judgments such as 'One act is *better* than an alternative because it has better consequences than the alternative.' A theory that makes only comparative judgments of better or worse between actions is a scalar morality.¹⁹ Clearly then, Slote's scalar morality is in all but name different from a continuum-valued morality designed to capture the fact that goodness and badness are scalar.

I said that my project would remain called for even if on my account being right is tantamount to being good. In point of fact however, I am not guilty of identifying rightness with goodness. For I hold that two acts are equally right iff they are performed for equivalent ends (foreseen consequences) and involve equivalent sacrifices. Thus, Bill Gates and I perform equally right acts if we each contribute our entire fortune to famine

¹⁸ Calculi of preference and their derivatives notwithstanding. See for example Georg Henrik von Wright's *The Logic of Preference*. Or Roderick Chisholm and Ernest Sosa's "Intrinsic Preferability and the Problem of Supererogation" and "On the Logic of 'Intrinsically Better.'"

¹⁹ *Common-sense Morality and Consequentialism*, Chapter V.

relief. Since Gates's contribution is bound to yield more goodness than mine, it follows that acts may be equally right and yet differ in the amount of goodness they produce. Conversely, if Gates and I each contribute \$500 to famine relief, our respective acts generate equal goodness, yet they differ in their degree of rightness. (A qualification is required. My act is more right than Gates's because it requires greater sacrifice. Since however sacrifice is nothing but a bad consequence, my act in fact produces less goodness than Gates's. To restore equivalence in goodness, let Gates's contribution be slightly lower.) This means that on my account, rightness and goodness are distinct properties.^{20 21}

9. Preliminaries: Descriptive and Normative Interests²²

I argued that rightness and wrongness are scalar properties and hence call for a many-valued morality, because (for example) rewards and punishments are proportionate to actions' degree of rightness and wrongness. This means that I infer formal or structural aspects of moral systems from common legal and moral practice. This in turn

²⁰ Interestingly, C. A. Campbell embraces an identical view in his *Selfhood and Godhood*:
 ...X is just as morally praiseworthy as Y or Z if he exerts an equivalent moral effort, even though he may not thereby achieve an equal success... (167. Italics his.)

In his *Morals by Agreement*, David Gauthier advances a very similar position with respect to rational (which for him is synonymous with 'just') co-operation. Specifically, Gauthier argues that just bargains require (in most cases) equal relative concessions from the partners, where 'relative concession' is roughly the ratio of benefit to investment (sacrifice). Thus, if *A* and *B* invest (respectively) \$40 and \$60 in a mutually profitable venture whose net benefit is \$200, their respective shares in the benefit should be \$80 and \$120 (each of them makes a relative concession of 2). Gauthier holds then that moral equivalence consists in equal *proportions* of benefit to sacrifice. In this respect his view resembles my own, since I too maintain that moral equivalence involves equal *proportions*—of wealth (ability) to sacrifice.

²¹ Incidentally, these comments also serve to reply to Alastair Norcross's "Scalar Approach to Morality." Norcross argues that consequentialism calls for a scalar notion of rightness; yet any notion of rightness which is both scalar and consequentialist entails that rightness is identical with goodness. (Supposing that Norcross is correct, I don't quite see why consequentialists should lose sleep over his conclusion.) On my account however, rightness is scalar, consequentialist (since the sacrifice involved in performing an act is part of its consequences, I in fact hold that rightness is a function of consequences alone), and differs from goodness.

²² This Section is greatly influenced by Robert Nozick's introductory remarks to his "Moral Complications and Moral Structures."

means that I am interested in how people *in fact* behave, not in how they *should* behave. There is then a sense in which my project is essentially descriptive.

There is however also a sense in which it is normative. For I am committed to the view that if a person's moral system is scalar, she ought to be troubled upon discovering that some of her positions conflict with this feature. Thus for example, a person who subscribes to many-valued morality should be alarmed to find out that she rejects mediation as means for conflict resolution.

This work then amounts to an attempt to strike an equilibrium between descriptive and normative interests in morality. In this respect it resembles contemporary decision theory, which reflects a continuous effort to strike a balance between descriptive adequacy and normative usefulness.

10. Summary

Punishments and rewards are proportionate to actions' *degree* of wrongness and rightness, which means that gradation is a datum about morality. I argued that this datum cannot be captured by existing—that is, two- and three-valued—moralities, but *can* be accommodated by a many-valued framework. In addition, many-valued morality has important ramifications for major issues in ethics. Assessing many-valued morality according to both the considerations that motivate it and its consequences for ethics naturally leads to the conclusion that such a system should be constructed.

Part I: Many-valued Morality

(As discussions in Chapter 1 suggest) the richer a system's expressive power (the higher the number of values it employs), the better it fares at accommodating moral gradation. Since however moral gradation forms a continuum (recall that I take rightness and wrongness to be a function of—*inter alia*—pleasure and pain: that is, of properties that presumably form continua), no *finitely*-valued framework is capable of capturing it *in its entirety*. A continuum-valued morality by contrast *is* capable of achieving that end (obviously). Constructing such a system—that is, a many-valued morality—is the aim of Part I.

I proceed by characterizing syntax and semantics (formation rules and truth-tables, respectively) for major moral connectives: namely, terms—such as 'not' and 'or'—that form a complex action description when attached to an action description or descriptions. Since the moral connectives bear close resemblance to the logical connectives, I conclude Part I with an examination of whether many-valued morality is akin to propositional logic also as far as semantic validity is concerned.

A couple of worries should perhaps be handled already at the outset (others will be more appropriately taken care of later on). In the first place, one might suggest that as a matter of fact there is no need for a continuum-valued morality, because in practice we employ only a limited variety of moral values. To this I reply that since unprecedented cases regularly make new demands on our expressive powers, we might as well—while we are at it—construct a maximally general framework and thereby anticipate those demands.

Another worry is that my talk of truth-tables (tables of *truth*-value) was a slip of the tongue; that I in fact meant to refer to tables of *moral* value. Well, the answer is that if the objection were correct, then—by parity of reasoning—there would be a need for baldness-tables, sweetness-tables, and so on. A need, that is to say, for a table for every scalar property. To avoid such massive violation of the methodological requirement of parsimony, it is customary to translate claims like ‘Herbert is bald to such-and-such degree’ into ‘“Herbert is bald’ is true to such-and-such degree.’’ Claims such as ‘Helping others when one’s mind is “overclouded by sorrows of his own” is right to such-and-such degree’ can likewise be translated into ‘“Helping others...is right’ is true to such-and-such degree.’’ (Contrary to what might be thought at first blush, such translation does not beg any major meta-ethical questions, as it doesn’t require a specific notion of truth: absolute and relative truth would serve equally well.) This remark notwithstanding, I shall often—in order to avoid cumbersome formulations—bracket truth-values and carry out the discussion in terms of moral values alone.

Chapter 2: 'Not'

1. Preliminaries (1)

Of the operations on actions that come to mind, negation is perhaps the most conspicuous. What exactly *is* negation of actions? To a first approximation, it amounts to *inaction* or non-occurrence of an action: the negation of nodding presumably consists in omitting a nodding or in failing to nod, and the negation of helping others—likewise—consists in failing to or refraining from help.

Whether helping others—that is, a *moral* action—does indeed take place may of course be hard to determine. For even if we for the moment focus our attention on a homogeneous moral community thus holding the variable of social or cultural conventions fixed, it's often unclear what constitutes such actions. Does giving a beggar (in Manhattan in 1999) a single cent constitute helping her? Definitely not. Giving a single cent under these circumstances amounts to *failing* to help. Giving her a single *dollar* on the other hand probably *does* constitute helping. And whether giving (say) five cents does so as well is simply unclear.

I bring up this matter not so much in order to point out that helping admits of degrees (no news to philosophers who—for a long while now—have recognized that very many properties are scalar). I bring it up rather for the purpose of stressing that admitting of degrees doesn't distinguish *moral* actions from *non-moral* ones. To use the same example as above, whether they constitute nodding may be unclear with respect to some head movements, yet perfectly clear with respect to others. This shows that it would be impertinent to object that, since it's sometimes difficult to tell whether moral

actions do indeed occur, their negation is very different from negation of non-moral actions; and therefore, the analogy which I draw here between the two types of actions in fact breaks down.

It would however be perfectly pertinent to complain that whether helping a friend occurs is not just a matter of what brought the action about—intentions—but also of what is brought about by it—objective consequences. Upon observing a small child attempting to help her parent with some chore, we may coherently and justifiably comment that she is not *at all* helping the parent (usually meaning that the child is actually interfering with the job). However, since I am assuming that objective consequences are morally irrelevant, I must stipulate—admittedly at the cost of betraying ordinary talk—that negation applies to only an intentional or conscious attempt to bring about a certain outcome. Not saving a life thus refers to consciously failing to make an attempt at—to refraining from—saving a life, not to an unsuccessful attempt to achieve that end.

2. Preliminaries (2)

The opinion that the distinction between action and (conscious) omission is morally relevant—that for instance killing differs from knowingly letting die—lies at the heart of a heated controversy.¹ Since the issue cannot be settled here (it requires a dissertation-length treatment), I shall provisionally assume that the distinction in question is morally significant (I *do* however discuss it at some length and defend it in Chapter 9, Sections 6-7 and Appendix 2). This assumption represents an almost wall-to-wall consensus—it is legitimate and called for in at least that sense. Besides, if it's

¹ See for example Chapter 3 of Shelly Kagan's *The Limits of Morality*. Or Patricia Smith's *Omission and Responsibility*.

false—if the distinction in question is morally *irrelevant*—negation is pointless and the talk of it meaningless.

A different challenge to the distinction between action and omission was leveled by Philippa Foot. Mrs. Foot has argued² that some actions—such as pulling the plug on an artificial respirator—should be treated as allowings, to the effect that the morally significant distinction is not between action and inaction, but rather between (i) initiating or sustaining a harmful causal sequence, and (ii) allowing or enabling a harmful causal sequence to run its course.³ This again is a controversy that cannot be taken up here.⁴ Nor *need* it be however: whether a certain action is more adequately treated as an allowing—and thus as a *negation* of an action—is a matter that must be settled *prior* to applying the formal apparatus I am about to offer. That this is so should be clear upon reflection.

3. Fixing Ideas: Right Actions (1)

So much for preliminary comments. Let us now turn to identification and formalization of (some of) the widely-shared intuitions that govern negation of right actions (*wrong* actions will be considered later on).

To fix ideas, suppose that an excellent swimmer attempts to save the life of a drowning child. Unbeknownst to the swimmer—or to anyone else for that matter—a deadly type of fungus flourishes in the water. An hour or so after coming in

² In a number of publications, among which is her “The Problem of Abortion and the Doctrine of Double Effect.”

³ While Foot’s interest is restricted to harm, her insights apply to benefit as well: one may initiate or sustain a *beneficial* causal sequence as well as allow such sequence to run its course. Imagine—to take an unsophisticated example—that food is *en route* to aid the starving inhabitants of some poor village and that Abdul refrains from robbing the caravan laden with the provisions. Then he is allowing a beneficial causal sequence to run its course.

⁴ A recent examination of it—and in particular of Warren Quinn’s stand on the matter—can be found in Samuel Rickless. “The Doctrine of Doing and Allowing.”

contact with it, both the swimmer and the child die. Now since (I am supposing) objective consequences are morally irrelevant, the swimmer's death doesn't detract from his action's degree of rightness.⁵ Does his death actually *add* to the action's degree of rightness? I think not. Not only because the swimmer did not *intend* to sacrifice his life, but also because had he known about the risk involved in his action, the latter would not have been right at all, but rather *very wrong*. At any rate, this question brings up an important issue—that of the relationship between an action's degree of rightness and the (foreseen) sacrifice involved in its performance—to which I now turn.

Saving a life at little risk or cost to oneself is *very* right, and refraining from so doing—*very very* wrong. Yet saving a life at a medium risk or cost to oneself is even *more* right, but (consciously) failing to do so is not *as* wrong. Fulfilling oneself—cultivating one's talents, if we are to employ Kantian terminology⁶—under circumstances of good health and relative affluence (what I have in mind specifically is the affluence level enjoyed by many inhabitants of Western countries) is certainly right, though not to an *unusual* degree; however, failing to do so under such circumstances—i.e., when life is so much on one's side—is *very* wrong. On the other hand, fulfilling oneself under (more or less) contrary circumstances—for example, given some sort of severe disability—is *unusually* right, yet failing to do so is perfectly understandable and hence only *mildly* wrong (but wrong nonetheless; it would seem that no circumstances

⁵ Being an excellent swimmer and unaware of the fungus, the swimmer's action *was* right. No doubt. Note however that if objective consequences are morally relevant, his action is wrong. For it results in *two* deaths (both of poisoning), whereas refraining from it would result in a *single* death (the child's, of drowning).

⁶ *Groundwork of the Metaphysic of Morals*, 90: 423. W. D. Ross employs the term 'self-improvement' (*The Right and the Good*, 21).

would render self-neglect entirely excusable).⁷ Finally, keeping a promise to buy groceries on one's way home is—under normal conditions—only slightly right, and knowingly failing to do so is only slightly wrong, though presumably the failure is more wrong than the performance is right.

One intuition that this series of examples brings to the fore is that the higher the foreseen sacrifice involved in performing a *right* action, the higher is its degree of rightness, and the lower the wrongness degree of failing to perform it.^{8 9}

4. Digression: Rightness, Sacrifice, and Virtue Ethics

Interestingly, the intuition that the harder a good action the higher its rightness degree presents a difficulty for virtue ethics (and I simply *must* add: *gey weis* what that is). This point is made by Philippa Foot in her “Virtues and Vices.” Virtue ethicists are both inclined and disinclined to think that the harder a person finds it to act virtuously the more virtue he demonstrates if he *does* act appropriately. For on the one hand great virtue is needed where it is especially hard to act virtuously, but on the other hand difficulty in acting virtuously indicates that the agent is lacking in virtue. It is thus

⁷ Some people may reject self-fulfillment as subject of moral evaluation. None of my subsequent reflections rest on this choice of example.

⁸ The former part of this intuition—the view that the higher the sacrifice involved in performing an action the higher the latter's degree of rightness—is the very idea that guides Kant in the early parts of his *Groundwork*. See Chapter 8.

⁹ In conversation, Professor Rohit Parikh argued that obligatoriness and rightness are independent properties, since actions may be equally right and yet differ in their degree of obligatoriness. I agree with Professor Parikh's claim on several occasions throughout this work. The reason that I mention the claim here is that the example which Professor Parikh contrived in order to establish it can be seen as yet another expression of the intuition that the higher the sacrifice involved in performing a good action the higher the action's degree of rightness. Following is the example. An attempt to save a drowning person made by an excellent swimmer is *very* right, and failing to do so *very very* wrong. What about an attempt to save a drowning person made by a lifeguard (on duty)? According to Professor Parikh, the latter attempt is as right as the former, but since it's more obligatory, refraining from it would be more wrong than refraining from the former case.

To me however it seems that the latter attempt is *less* right than the former. This is because the lifeguard has entered a contract, which makes attempting to save drowning people his duty, and whose breach is tied to certain sanctions. These sanctions serve as a deterrent against refraining from attempting to save drowning people, thus decreasing the effort involved in attempting to save the drowning person.

unclear who shows most courage, the one who wants to run away but does not, or the one who does not even want to run away.¹⁰

Is it also unclear from my point of view? My stand on the issue is of course that the agent who feels fear but overcomes it shows more courage (and if he fled, he would be held less blameworthy). His very experience of fear doesn't indicate that he is not predisposed to behave courageously; rather, it's a *precondition* for courage.

5. Fixing Ideas: Right Actions (2)

I said that one intuition that the examples of Section 3 bring to the fore is that the higher the foreseen sacrifice involved in performing a *right* action, the higher is its degree of rightness, and the lower the wrongness degree of failing to perform it. Another intuition that these examples elicit is that the more worthy the cause for which an action is performed (i.e., the higher the foreseen worth of the action's goal), the higher is its degree of rightness.

These intuitions interact in a rather interesting way. If an action is such that performing it involves little effort or risk and its aim is not particularly worthy—an example would be keeping a promise to buy groceries on one's way home—then its degree of rightness is slight, and failing to perform it would be only slightly wrong. If an action involves little sacrifice, but its aim is worthy—saving the life of a Somali child by contributing a small amount of money would be a good example—then performing it is *very* right, yet failing to perform it *very very* wrong. If an action involves considerable sacrifice, and its goal is not particularly worthwhile—an example would be putting one's life at a considerable risk in order to save an object, precious though it may be, from a burning house—then performing it is *very* wrong while failing to perform it

¹⁰ 589.

very right. Finally, if an action requires much sacrifice, and its goal is extraordinarily worthy—as for example risking one’s life, or even sacrificing it, in order to save very many lives—then performing it is *very very* right, while failing to perform it is only *mildly* wrong (though not, it would seem, entirely devoid of wrongness. I expect at least some readers to find this opinion troubling. I touch upon it in Chapter 9.) (Clearly however if the failure amounts in effect to evading compulsory service in the military when such service is a burden that every member of one’s community shares in their turn, then the failure is *very very* wrong.)

6. Truth-table (1)

We are interested then in four kinds of actions, which can be conveniently (though crudely) represented by a 2x2 table in the following way.

Sacrifice	Negligible	Considerable
Cause's Worth		
Low	Slightly right	Wrong
	a	b
High	Very right	Very very right
	c	d

Fig. 1.1. ‘Not’: Right Actions

Transitions from one kind of action to another are of course gradual (which is why the representation is crude). It’s unclear for example whether sacrificing one’s life in order to save a newborn baby is right or wrong (i.e., whether it falls into cell (d) or into cell (b) instead). Nor is it clear how many lives must be at risk in order for one’s failure to

sacrifice oneself (in order to save these lives) to qualify as a wrongdoing: after all, it stands to reason that failing to sacrifice one's life in order to save two or three lives is not a wrongdoing. Such issues should—like relevantly similar questions about utility—be addressed *empirically* by social and medical scientists.^{11 12}

At any rate, what *is* clear is that if an action is right to a certain degree, then failing to perform it is *not* wrong to the *same* degree: fulfilling oneself under conditions of good health and relative affluence is *mildly* right, but failing to do so is *very* wrong. Similarly, supporting a Somali child (at a monthly cost of \$20, the real cost to date) is *very* right, but failing to do so is *very very* wrong.¹³ The relation then between right actions and their negations is thus not a simple one.

A pattern does nonetheless begin to emerge. Let the value +.5 be assigned to right actions whose performance involves a negligible sacrifice and serves a very worthy cause. Then failing to perform such actions, though not *very* close to -1 (these values are reserved to actions like premeditated murder; I shall say more about this matter shortly), is *rather* close to it. Actions whose performance involves a considerable

¹¹ I elaborate on this remark in Chapter 9.

¹² Although inquiries into issues of this sort may seem to some readers—at least at first glance—rather remote or abstract, they are in fact of direct concern and relevance to us all (almost). This is because we tend to reside in communities that keep a defense force, a fact which requires each of us to agree on the conditions under which they would be willing to defend their community: that is, to put their lives at a considerable risk in order to save the lives of other community members.

¹³ Indeed, I look upon such support not as a supererogatory act the failure of whose performance is not at all wrong, but rather as an obligatory act the failure of whose performance is wrong. In this respect my view converges with that of, among others, Peter Singer (“Famine, Affluence and Morality”) and Shelly Kagan (*The Limits of Morality*).

By subscribing to this view I do not mean to suggest (nor do I *actually* suggest) that the supererogatory should be an empty category. Maintaining it—as well as the counterpart category of the suberogatory (wrong but not forbidden actions)—populated is of utmost importance. For if all right actions were obligatory and all wrong actions forbidden, persons would be left with little or no room for pursuing projects such as achieving an intelligent understanding of the world; that is, projects of the sort that tends to provide (at least some) persons with a reason for living at all. (The influence of Bernard Williams’s “Persons, Character and Morality” is perhaps obvious.)

sacrifice and a very worthy cause would have a rightness degree higher than +.5, but refraining from them would be much less than maximally wrong. Actions whose performance involves little sacrifice, and whose goals are not particularly worthy would enjoy a rightness value lower than +.5; failing to perform them would have a low wrongness value (lower than -.5, let us say). It accordingly seems appropriate to stipulate that up to .5, the wrongness degree of failing to perform a right action is -1.7 times its degree of rightness. From this point onward, the wrongness degrees mirror previous wrongness degrees. The resulting truth-table for negation is presented below (A_R stands for a right action, and $\neg A_R$ for refraining from A_R).

A_R	$\neg A_R$
.1	-.17
.2	-.34
.3	-.51
.4	-.68
.5	-.85
.6	-.68
.7	-.51
.8	-.34
.9	-.17
.99	-.017

This table can be cast more compactly in the following way, where v is an action's degree of rightness ($v[A_R] > 0$):

$$\begin{aligned} \text{If } v[A_R] \leq .5, v[\neg A_R] &= -1.7 v[A_R] \\ \text{If } v[A_R] > .5, v[\neg A_R] &= -1.7(1 - v[A_R]) \end{aligned}$$

7. C'est La Vie

I had hoped to reach a *comprehensive* systematization of the foregoing considerations. To reach, that is to say, formulae for calculating not merely the negation value of actions, but also an action's moral value from (i) the worth of its cause and (ii)

the value of the sacrifice involved in performing it. But no (non *ad hoc*) such formulae are forthcoming. I fear.

Specifically. We are looking in effect for a computation procedure that takes the values on the left as input and yields the values on the right as output. (The numerical values are meant to be representative examples.)

	Input		Output
	Cause's value	Sacrifice's value	Action's moral value
1.	high (.9)	high (.9)	.9
2.	high (.9)	low (.1)	.5
3.	low (.1)	high (.9)	negative
4.	low (.1)	low (.1)	.1

Averaging the input values is a reasonable path to take, plus it yields the desired output (try that out) in rows 1, 2, and 4. Not so in row 3 however (try *that* out too, if there's a need). In *this* row, subtracting the value of the sacrifice from that of the cause yields the desired output (-.8). Now it is of course possible to stipulate that (say) whenever the value of the cause is lower—and the value of the sacrifice higher—than .5, the action's moral value equals the difference between the former and the latter. But such a move clearly calls for a justification. The latter is alas difficult to come by. And its absence renders the stipulation in question intolerably *ad hoc*. This result is naturally a disappointment for me. But such is life.

8. -1.7

Returning now to the truth-table for negation, it may seem unreasonable (or at least aesthetically obtuse) to select -1.7—rather than the round -2—as the negation coefficient, as the latter would produce a smaller number of ugly negation values.

I answer that if the negation coefficient were -2, the negation of an action that enjoys .5 degrees of rightness would be -1, which is an undesirable result. For one thing, -1 and 1 may not be assigned to acts within the scope of this system: I explain why in Section 15. Secondly, if the negation value of an act enjoying .5 degrees of rightness were -1, it would imply that refraining from actions that involve a negligible sacrifice and a very worthy cause is *as wrong as* premeditated murder. And this seems counter-intuitive: knowingly failing to prevent the loss of a life at practically no cost—say, refraining from supporting a Somali child—is not *as wrong as* actively terminating a life, or else we would all be murderers. For this reason I said earlier that refraining from right actions that involve a negligible sacrifice and a worthy cause is only *rather* close to being of value -1, not *very* close to it.

(I said that if letting die is as wrong as killing, we would all be murderers. I should like to note that this is in fact a *bad* argument for the view that the (positive) duty to help others in need is less stringent than the (negative) duty to forbear from harming others. And it's bad because—like so many other arguments for that view—it presupposes the very position that it's designed to support. Here is why. The argument under consideration can be cast as follows.

1. If refraining from helping others is as wrong as actively harming them, then we are all murderers. For we all knowingly fail to support starving children in Somalia (even though such support requires only a negligible sacrifice on our part).
2. Surely however we are not all murderers.
3. Therefore, refraining from helping others is not as wrong as actively harming them.

But why think that we are not all murderers (that premise 2 is correct, that is)? The answer presumably is that we do not all actively cause deaths; we merely forbear to

prevent them. Further instances of circular arguments for the view that doing harm differs from merely allowing it appear in Chapter 9, Appendix 2.

As I said in Chapter 1 and at the outset of this Chapter however, I here *assume* that the distinction between action and inaction is morally significant.)

A second remark concerning the negation coefficient is perhaps somewhat superfluous, but—to be on the safe side—I shall make it nonetheless. –1.7—unlike for instance the speed of light and other natural constants—is not an objective magnitude. Rather, it’s a magnitude that yields negation values which cohere well (or even best) with certain major moral intuitions.¹⁴

9. Supererogation

The truth-table for negation of right actions may not yet be complete. This is because it doesn’t include supererogatory acts: that is, right actions the failure to perform which is not at all wrong,¹⁵ so that their negations are invariably of value 0. (This is so even though supererogation admits of degrees: doing someone a small favor is—by definition—supererogatory, but sacrificing one’s life in order to save three or four lives is heroic. Yet failing to perform either act is presumably not at all wrong, and therefore of value 0.) To complete the truth-table, the following line needs to be added to it.

¹⁴ See David Lewis’s “How to Define Theoretical Terms.”

¹⁵ As David Heyd maintains in his *Supererogation*, this definition of the supererogatory is defective. For its positive term fails to refer to the *value* of the acts beyond duty (117), which means that it doesn’t exclude actions that realize (say) human excellence: that is, it doesn’t exclude a Nietzschean type of morality (134). To correct this defect, it must be required that the act realizes a *moral* value (134; see also 115).

A_R	$\neg A_R$
≥ 1 ¹⁶	0

Now according to this line all supererogatory acts are of value 1 or higher, which means that any supererogatory act is more right than any obligatory act. Is this view correct? In her "The Supererogatory" Professor Julia Driver argues that it cannot be.^{17 18} Suppose that I have contracted to clean your house and am therefore obligated to do so. If I then clean both your house and your garage, I have performed in excess of this obligation: the cleaning of the garage is supererogatory. Compare my conduct with a captured soldier who is being tortured and who nevertheless refrains from revealing the location of his comrades, thus doing what he has an obligation to do. If the view under consideration is correct, my conduct has a higher moral value than the soldier's. But this cannot possibly be so: the soldier's act is certainly more praiseworthy than the supererogatory act of cleaning a garage.¹⁹

Not every supererogatory act then is more right than any given obligatory act.²⁰

Thus, the line

¹⁶ Recall that if $v[A] > .5$, $v[\neg A] = -1.7(1-v[A])$. Therefore, if $v[A] = 1$, $v[\neg A] = -1.7(1-1) = 0$.

¹⁷ Professor Driver employs in fact not 'right' but rather 'praiseworthy.' The two terms however diverge only rarely. For further discussion of this matter, see Appendix to this Chapter.

¹⁸ Roderick Chisholm offers an argument to the same effect ("Supererogation and Offence: A Conceptual Scheme for Ethics," 3-4).

¹⁹ 290. Professor Driver suggests that people generally assume that supererogatory acts are always more praiseworthy than obligatory acts, because they mistakenly think of supererogatory acts as being extraordinary and heroic in nature (289-90).

²⁰ However, as Frances Kamm argues in "Supererogation and Obligation," since some supererogatory acts *are* more right than some duties, a supererogatory act may take precedence over a duty. To see this, imagine that I have promised to meet someone for lunch, and am thus under an obligation to keep my promise. On my way to the lunch appointment I come across a car crash. One of the victims needs a kidney transplant in order to survive, and I am the only person available who has the right kind of kidney. Given the size of the sacrifice to me, it is not my duty to donate my kidney. In these circumstances then the only *duty* that I have is to keep my promise.

Suppose however that I am willing to donate a kidney, and that going through the required medical procedure would cause me to miss my appointment. Then even though keeping my promise is still my only duty, I may fail to do it in order to donate a kidney (and thus save a life) (119-20).

A_R	$\neg A_R$
≥ 1	0

must *not* be included in the truth-table for negation.

One might wonder: if the trouble lies with the position that every supererogatory act is more right than any obligatory act, why not simply abandon it? The answer is that doing so means in effect that two actions with the same rightness value may have different negation values: for if one of these acts is supererogatory and the other obligatory, the negation value of the former would be 0 whereas that of the latter negative. Negation would thus assign to at least some members of its domain more than a single member of its range, thereby failing to be a *function*. We must therefore give up hope of integrating the supererogatory into the formal framework under construction.²¹ As we shall see later on, the same is true of the suberogatory: namely, wrong but not forbidden actions.

10. Fixing Ideas: Wrong Actions

So far I have considered negations of right actions only. What can be said about negations of *wrong* actions? Remarks in the same spirit as those that were made concerning negations of right actions, as a matter of fact.

Begin by noting that although premeditated murder is (presumably) maximally wrong,²² a theory on which consciously refraining from such action is *maximally* right

²¹ Waiving the requirement that evaluation be a *function* is also an alternative of course. That is, one may opt for a system in which—as is the case with paraconsistent logics—evaluation is not a function but instead a *relation* that sometimes assigns to one and the same member of the domain more than a single member of the range. I intend to explore this possibility elsewhere.

²² I employ ‘maximally wrong (right)’ for convenience only. Strictly speaking, there are no *maximally* wrong and right actions. For as I explain in Section 15, no obligatory or forbidden act—that is, no act within the scope of this system—may be assigned the values 1 and -1.

would be very disagreeable indeed. Intentionally avoiding maximally wrong actions is surely right,²³ but not to an unusual degree: *such* values ought to be reserved for actions that involve considerable sacrifice and a worthy cause. Put in slightly other words (and taking an example), refraining from murder is not as right as risking one's life to save two or three lives. It follows that just as the negation of a maximally *right* action is not maximally wrong, the negation of a maximally *wrong* action is not maximally right.

What then would be an example of an action refraining from which is close to being maximally right? Murdering a wealthy person for his fortune, given that one is destitute and the chances of being caught are practically nil is presumably such an action.²⁴ In general, refraining from a wrong action is close to being of value 1 if some of the action's foreseen consequences are very bad, while others—usually those that the agent foresees for himself—are so good as to make the action almost irresistibly tempting. (Indeed, the foreseen good consequences of an action are not distinct from the temptation to perform it, but are instead *constituents* of that temptation.) So much—for the time being—for actions whose performance is highly tempting but heavily costly (for the victim(s)).

²³ I *do* think that *unintentional* failure to murder or cheat is morally neutral: if it were right, then we would all—by the sheer fact of being alive—be performing indefinitely many right actions (I am in agreement on this matter with the position evinced in Roderick Chisholm and Ernest Sosa, "Intrinsic Preferability and the Problem of Supererogation," 329). However, in his *Ethics and the Limits of Philosophy* Bernard Williams makes a point worth repeating in this context:

An effective way for actions to be ruled out is that they never come into thought at all, and this is often the best way. One does not feel easy with the man who in the course of a discussion of how to deal with political or business rivals says, "Of course, we could have them killed, but we should lay that aside right from the beginning." It should never have come into his hands to be laid aside. ...[S]ome concerns are best embodied in this way, in deliberative silence (185).

²⁴ On this position too I am in agreement with Chisholm and Sosa's "Intrinsic Preferability and the Problem of Supererogation."

If the temptation to do a wrong is very high, but the cost (to others) is not particularly heavy—an example would be a state of affairs where one can hardly resist stealing a comparatively worthless object from a well-to-do person—then failing to perform the act would not be of an unusually high degree of rightness. If the temptation is rather low and the cost high—murdering a poor person towards whom one feels no special hostility would be a good example—then refraining from the action isn't particularly right. (That is, if no special effort is required to refrain from a very wrong action, refraining from it is not unusually right.) Finally, if both temptation and cost are low—as is the case when (say) one is not especially tempted to steal a relatively worthless object—then avoiding the action is of only a low degree of rightness. In conclusion, one principle governing negation of wrong actions is that the less the temptation to perform a wrong action—the lower the effort required to refrain from the action—the lower the rightness degree of refraining from it.

Now in many cases the overall temptation to do a wrong is influenced by the action's expected costs not only to the agent, but also to others. If Cohen is desperately poor and the chances of him being caught are small, he may find stealing the life savings of a working class family much less tempting than stealing the same amount of money from a wealthy family (because his act is likely to bring much greater misery to the poor family). The principle at hand implies that failing to steal from the wealthy family is *more* right than failing to steal from the working class family. Although this result is without a doubt unwelcome, there are considerations in light of which we should be willing to bite the bullet.

The circumstances which constitute the temptation to do a wrong also affect our assessment of the agent's degree of responsibility for the act, thus affecting the latter's degree of wrongness. If the agent grew up in an underprivileged environment, he would in all likelihood be very tempted to steal; and if he *were* to steal, his life circumstances are likely to be looked upon as mitigating circumstances as well. If on the other hand the agent grew up in an affluent surrounding, he is likely to be less tempted to steal; and if he were to steal, his life circumstances would be regarded as aggravating circumstances too. This means that there is a second principle that wrong actions follow. It is that the greater the temptation to perform a wrong action, the lower its degree of wrongness.

Taken together, the two principles state that the bigger the effort required to *refrain* from a wrong action, the *lower* the action's degree of wrongness and the higher the rightness degree of refraining from it. To use the same example as above, if Levi—against all odds (created by his life circumstances)—refrains from stealing, his behavior is *very* right; but if he were to steal, his action would not be *that* wrong. It is the plausibility of the first principle (according to which the smaller the effort required to refrain from a wrong action, the lower the rightness degree of refraining from it) in conjunction with the second principle (which states that the greater the temptation to perform a wrong action, the lower its degree of wrongness) and in contexts like the one at hand which explains why we should be willing to bite the bullet associated with the first principle.

11. Truth-table (2)

By now it's clear that, as with right actions, with wrong actions too we are interested in four types of action (the transitions between which are gradual). They can be represented by a 2x2 table in the following way.

Temptation	Low	High
Cost		
Low	Wrong a	Slightly wrong (right?) b
High	Very very wrong c	Very very wrong, but not as in cell c d

Fig. 1.2. 'Not': Wrong Actions

As far as right actions are concerned, the greater the effort required to *perform* an action, the *higher* its degree of rightness and the *lower* the degree of wrongness of avoiding it. As far as wrong actions are concerned, the greater the effort required to *refrain* from an action, the *lower* its wrongness degree and the *higher* the rightness degree of refraining from it.

In accordance with the foregoing observations, let -.5 be the value of wrong actions whose performance is both overwhelmingly tempting and highly costly (for the victim(s)). Refraining from such actions would be rather close to being maximally right (that is, to being of value 1). Actions whose performance is very tempting and not particularly costly have a wrongness degree lower than -.5 (though a (consequentialist) case could be made for taking such actions to be right: after all, omitting the action is costly yet committing it isn't); failing to perform them would be only slightly right (or—

if they are considered right—only slightly wrong). Actions whose performance is not particularly tempting and whose goals are rather worthless would have a wrongness degree lower than $-.5$; avoiding them would have a low rightness value (lower than $+.5$, let us say).²⁵ It seems appropriate to stipulate that up to $-.5$, the rightness degree of refraining from a wrong action is -1.7 times its degree of wrongness. From this point onward, the rightness degrees mirror previous rightness degrees. The resulting truth-table for negation is presented below (A_w stands for a wrong action, and $\neg A_w$ for knowingly or consciously failing to perform A_w).

A_w	$\neg A_w$
-.1	.17
-.2	.34
-.3	.51
-.4	.68
-.5	.85
-.6	.68
-.7	.51
-.8	.34
-.9	.17
-.99	.017

More succinctly:

$$\begin{aligned} \text{If } v[A_w] \geq -.5, v[\neg A_w] &= -1.7 v[A_w] \\ \text{If } v[A_w] < -.5, v[\neg A_w] &= -1.7 (1+v[A_w]) \end{aligned}$$

The two truth-tables for negation can be unified in a single table:

$$\begin{aligned} \text{If } -.5 \leq v[A] \leq .5, v[\neg A] &= -1.7 v[A] \\ \text{If } v[A] > .5, v[\neg A] &= -1.7(1-v[A]) \\ \text{If } v[A] < -.5, v[\neg A] &= -1.7(1+v[A]) \end{aligned}$$

²⁵ Recall that examples of each kind of action appear in Section 10.

Let us represent this truth-table on an axes system, taking the points on the x axis to stand for values of actions, and the points on the y axis to stand for values of negations of actions.

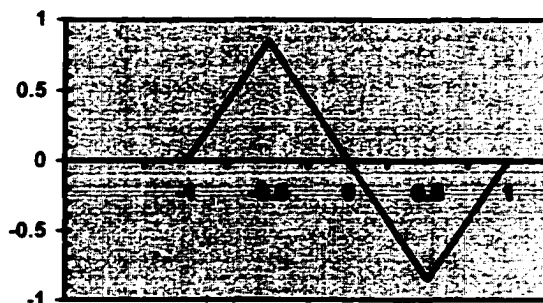


Fig. 1.3. 'Not': Graphic Representation

12. Further Disappointment

In connection with wrong actions too I had hoped to offer formulae for computing an action's moral value, this time from (i) its cost and (ii) the temptation it poses. But this case resists (non *ad hoc*) systematization to an even greater extent than its counterpart in Section 7. A gaze at the chart should make this evident. (You will recall that we are searching for a computation procedure that takes the values on the left as input and yields the values on the right as output. As before, the numerical values are intended to be representative examples.)

	Cost	Input Temptation	Output Action's moral value
1.	high (.9 or .8)	high (.9 or .8)	-.5
2.	high (.9 or .8)	low (.1 or .2)	-.9
3.	low (.1 or .2)	high (.9 or .8)	-.1 (positive?)
4.	low (.1 or .2)	low (.1 or .2)	-.1

13. Cost and Cause, Temptation and Sacrifice

The value of a wrong action is determined by two factors. The first is the action's cost—the badness of its foreseen consequences (for others usually). The second is the temptation to commit the action: namely, the foreseen effort or sacrifice (for the agent) required for refraining from it.

The value of *right* actions too has two determinants. The first is the foreseen worth of the cause for which the action is performed: that is, the goodness of its foreseen outcomes (usually for others). The second is the value of the foreseen sacrifice (for the agent) involved in its performance, which can presumably be understood in terms of the temptation to refrain from the action (as the sacrifice involved in performing an action increases, so does the temptation to refrain from it).

Arguably then, right and wrong actions are determined by exactly the same pair of factors.

14. Suberogation

Return now to the truth-table for negation of wrong actions. One might think that—like the previous table—it too is incomplete, this time because it does not include not *supererogatory*—but rather *suberogatory*—actions.²⁶ That is, actions whose commission is wrong but not forbidden. To take an example, suppose that *A* knows concerning *B*, whom *A* dislikes, that the loss of *B*'s job would result in great tragedy for *B* and his family; that there is another person, *C*, who could do *B*'s work but no more satisfactorily than *B* does it; and that *B*'s employer, even if he knew the foregoing, would replace *B* by *C*, if he thought that *C* were available. It is plausible to maintain

²⁶ Chisholm dubs suberogatory actions 'offences' ("Supererogation and Offence").

that if *A* were deliberately to bring the availability of *C* to the attention of *B*'s employer, it would be permissible but at the same time heinous.²⁷ (This would be so, I should think, even if *A* has justifiable reasons for disliking *B*.) To take another example—this time from the real world—under the existing law, a company can deprive its shareholders of their rights by establishing a subsidiary company and transferring to it the property of the mother-company. Although such a move is possible because of what is at least sometimes perceived as a loophole in the legal system, as long as the latter remains in place, the former is wrong but legally permissible. (It is of course true of any loophole in the law that exploiting it is wrong but permissible.)

Now according to Chisholm, failing to perform a suberogatory act—just like failing to perform a supererogatory act—is neither right nor wrong, and thus of value 0.²⁸ All this suggests that the truth-table for negation of wrong actions should include the following line:

A_w	$\neg A_w$
$-1 \geq$ ²⁹	0

But it shouldn't. The reason is familiar already. Just as not all supererogatory acts are heroic or saintly, not all suberogatory acts are heinous or diabolical: suberogation—like supererogation—admits of degrees. This means that at least some suberogatory acts are less wrong than some forbidden ones.³⁰ This in turn means that

²⁷ The example is Chisholm's (*ibid.*, 5).

²⁸ *Ibid.*, 10-1.

²⁹ If $v[A] < -.5$, $v[\neg A] = -1.7(1+v[A])$. Thus, if $v[A] = -1$, $v[\neg A] = -1.7(-1 + 1) = 0$.

³⁰ In fact, it seems to me that there are by far fewer suberogatory actions which are more wrong than forbidden ones than there are supererogatory actions which are more right than obligatory ones.

Professor Michael Levin suggested that *no* suberogatory act is more wrong than any forbidden one. I disagree: smoking in restaurants is forbidden, riding a motorcycle isn't; and the latter is more wrong than the former.

two actions with the same wrongness value may have different negation values. For if one of these acts is suberogatory and the other forbidden, then the negation of the former would be of value 0 whereas that of the latter of some positive value. Thus, negation would assign to at least some members of its domain more than a single member of its range, thereby failing to be a *function*.³¹ It must be conceded therefore that the suberogatory, just like the supererogatory, cannot be captured by the formal system under construction.

15. 1 and -1

Acts within the system's scope may be assigned any real number between -1 and 1. The negation value of 1 and -1 is 0. (Recall that if $v[A] = 1$, $v[\neg A] = -1.7(1-1) = 0$. Likewise, if $v[A] = -1$, $v[\neg A] = -1.7(-1 + 1) = 0$.) Any act of value other than 0 whose negation value is 0 is—by definition—super- or suberogatory. As we have seen, such acts fall outside the scope of the present framework. This means that no act within its scope may be assigned 1 or -1. Indeed, it means that the system's interval is open.

16. Double Negation

Failing to perform an action of value -.1 enjoys .17 degrees of rightness. Yet refraining from an action whose rightness degree is .17 is *not* of value -.1, but of value -.289 (.17 times -1.7) instead. Likewise, the negation of .7 is -.51, whereas the negation of -.51 is *not* .7 but rather .833. Letting $A \Leftrightarrow B$ iff $v[A] = v[B]$ (where A and B are any two formulae), this means that $\neg\neg A \Leftrightarrow A$ fails in my system.³² This in turn means that refraining from refraining from an action is—*morally* (not metaphysically, that is)—

³¹ But see fn. 21.

³² As it does in intuitionist logic.

unlike committing it. Refraining from refraining from murdering my wealthy neighbor Yankel differs morally from murdering him, and refraining from refraining from donating money to famine relief differs from donating the funds.

Is this consequence of the foregoing observations and stipulations palatable? We need some numerical examples before our eyes.

- (1) Saving the life of a starving Somali at little monetary cost enjoys .5 degrees of rightness. Refraining from such action is of value $-.85$. Refraining from refraining from the action enjoys $.255$ ($-.15$ times -1.7) rightness degrees.
- (2) Murdering Yankel when the temptation to do so is small is of value $-.9$. Refraining from it is of value $.17$ ($-.1$ times -1.7). And refraining from refraining from the action is of value $-.289$.
- (3) Rescuing three lives at risk of one's own life is of $.9$ degrees of rightness. Refraining from so doing is of $.17$ wrongness degrees (of value $-.17$, that is to say). Refraining from refraining from the action enjoys $.289$ degrees of rightness.

While the first example is merely perplexing (why should the eventual act enjoy merely half the rightness enjoyed by the initial one?), the second and third are all but intolerable (why should the eventual murder of Yankel be of a relatively low degree of wrongness? After all, the temptation to perform the act is small this time round as well. And why should the eventual act in the third example—which involves a risk of no less than the agent's *life*—enjoy only a relatively low degree of rightness?). These examples suggest then that the failure of double negation is not simply an unpalatable—but actually an unbearable—consequence.

They also suggest however that the failure may not be a consequence in the first place. Consider example (1). What story could be told to breath life into it? Perhaps this. The agent recognizes that she should make a negligible sacrifice in order to save the life of a starving Somali and at first refrains from so doing, but later she refrains from refraining from the action. Now this story leaves much in the dark (needless to say). Why does the agent first refrain? Perhaps because she is a procrastinator. But if this is so, then strictly speaking she does not refrain from the action but merely postpones its performance. Or perhaps she first refrains because she would rather spend the funds on a birthday present for her sweet niece. If *that's* the case, then the sacrifice involved in performing the action is by no means *negligible* for the agent.

Next turn to this question. When does the agent refrain from refraining?

Suppose that an agent resolves to go on a diet but refrains from so doing. A year later she finally goes on a diet. Does she refrain from refraining? If she does, her initial refraining spans over a year. But this is highly implausible: *refraining* from an action (consciously avoiding it) can presumably last only so much time. Finally, how come the agent later refrains from refraining? If it's because she can now afford the donation, then—again—the sacrifice involved in the initial act is not negligible for her, contrary to our assumption.

In short, a case could be made to the effect that we don't *understand* what refraining from refraining *means*. Or at least to the effect that it has no clear established sense distinguishable from that of more natural expressions such as performance or

commission.³³ This paves the way for a solution to the problem at hand: disqualifying double (and triple and so on) negation as a well formed formula (wffs) (that is, giving up the recursive nature of the formation rule for negation). The idea is not that only understandable formulae—formulae for which we have a general method of decoding—qualify as well formed ones (this is not a *bad* or uncommon idea, but I don't wish to delve into it here). It is rather that giving up double (and triple, and so on) negation exacts a relatively low price due to the fact that it's in any event useless (we don't understand what it means).

In conclusion, short of rejecting the principles and stipulations that result in failure of double negation, two avenues may be pursued. The first is justifying the consequences of the failure in question, and the second—banning double negation from the circle of wffs. The first route requires costly acrobatics. The second on the other hand is defensible (if somewhat *ad hoc*).

17. Ari & Son

In the previous Section I in effect argued that since realistic examples of refraining from refraining from acts are difficult (perhaps even impossible) to come by, we don't understand what 'refraining from refraining from an act' *means*. In response, Professor Michael Levin contrived Ari & Son.

Ari should, and knows he should, send his son at college money. But he knows his emotions often tempt him to do the wrong thing; in particular, pictures of starving

³³ This conclusion is reminiscent of Dorothy Edgington's discussion of some embedded conditionals, such as conditionals in antecedents of other conditionals and negated conditionals. Edgington points out that some such constructions—for example, 'If Kripke was there if Strawson was, then Anscombe was there'—cannot be deciphered. Other constructions—for example, 'It's not the case that if it rains, they will be delayed'—have no clear established meaning distinguishable from 'If it rains, they won't be delayed' ("On Conditionals," 283-4).

Somalis may make him refrain from sending his son money, and send it to Somalia instead. To keep himself from refraining from sending the money, he therefore avoids magazine articles about starving Somalis. He stops himself from stopping himself from sending the check.

Professor Levin argues that Ari & Son demonstrates the *plausibility* of distinguishing (i) the moral value of Ari's sending the money as the result of stopping himself from sending it elsewhere from (ii) the value of his just sending it as a matter of course. Let's assume (Professor Levin continues) that sending the money as a matter of course has a value of .6 (very worthy cause, small sacrifice). His failing to refrain from doing it has the slightly lower value .544.³⁴

I agree that Ari's sending the cash as a matter of course differs morally from his sending it as a result of stopping himself from sending it elsewhere. After all, the latter involves a greater sacrifice on his part. This however means that the latter act should enjoy not a lower, but rather a *higher* value than the former. Thus, Ari & Son—like the series of examples I presented in Section 16—at best serves to demonstrate that the failure of $\neg\neg A \Leftrightarrow A$ has perplexing or intolerable ramifications. 'At best,' because in order for Ari & Son to demonstrate that the failure of $\neg\neg A \Leftrightarrow A$ has perplexing consequences, Ari's sending the funds as a result of stopping himself from sending it elsewhere must be identical to his *refraining from refraining* from sending the check to Son: that is, it must be identical to $\neg\neg A$. But it isn't. Clarification follows.

One might complain that—not unlike Descartes's 'I don't exist'—the foregoing discussion is self-refuting. For contrary to its declared purpose, its very occurrence

³⁴ $-1.7 \times (1 - .6) = -.68$, and $-1.7 \times (1 - .68) = .544$.

proves that talk of refraining from refraining from acts is perfectly intelligible. Or that we *do* understand what 'refraining from refraining from acts' means. I disagree. Ari & Son establishes only that we understand what it means to perform one act while overcoming a temptation to carry out *another* (to act in accordance with duty and contrary to one's inclination, if you will). Ari doesn't refrain from refraining from sending money to Son; rather, he refrains from sending the funds to starving Somalis (by refraining from reading magazine articles on them), and instead sends it to Son.

Thus, to accommodate the intuition that Ari's sending the funds as a matter of course differs morally from his sending it as a result of stopping himself from sending it elsewhere, there is no need to admit $\neg\neg A$ as a wff.

18. ± 0.5

Turn now to an objection concerning stipulations such as that .5 is the value of right actions whose performance involves a negligible sacrifice and serves a very worthy cause. Or that -.5 is the value of wrong actions whose performance is both overwhelmingly tempting and highly costly (for the victim(s)). (In order to avoid unnecessary repetitions, I shall from now on confine my discussion to right actions: everything that I shall say about them equally applies—*mutatis mutandis*—to wrong actions.)

The worry is that there may be counter-examples to these stipulations: for instance, actions of value .5 whose performance serves a worthy cause and involves *considerable* sacrifice. Since the negation value of such acts equals -.85, it follows that their omissions don't enjoy only a low wrongness value (a value higher than -.5). This result conflicts with another stipulation of mine: namely, that negations of acts that serve

a worthy cause and involve a significant sacrifice are of a low wrongness value. So if counter-examples of the sort under consideration can be generated, the present formal framework would be in disarray.

Is it possible to generate counter-examples of this sort? If it *is*, the way to go about it is presumably the following. We first suppose that a non-moral action—namely, an action outside the moral realm whose moral value is 0—requires a considerable sacrifice. Then we hold the sacrifice fixed, and gradually enter the moral realm. Since the transition from the non-moral to the moral realm is gradual (whether an action is within morality or outside it is presumably a matter of degree), we shall at a certain point strike an action that involves considerable sacrifice and enjoys .5 degrees of rightness. Whether or not this action is performed for a worthy cause need not concern us, because as long as it's of value .5 and involves considerable effort it serves as a counter-example to the stipulation that every action of value .5 involves a negligible sacrifice and a worthy cause. (However, if the action *is* performed for a worthy cause, it constitutes a counter-example to the stipulation that every action which involves a considerable sacrifice and a worthy cause is of a moral value greater than .5. And if the action is done for an unworthy cause, it serves as a counter-example to the stipulation that every action which involves a considerable sacrifice and an unworthy cause is actually *wrong*.)

There is however a major difficulty for this objection; it concerns the initial supposition that there exists a non-moral action which involves considerable sacrifice. This supposition presumably means that the action in question requires for example putting one's life at great risk for the sake of some non-moral cause such as riding a

motorcycle. However, the very fact that the action requires risking one's life qualifies it as moral, and the very fact that it's performed for the sake of a non-moral cause qualifies it as morally *wrong*.³⁵

This reply is admittedly *ad hoc* and does not show that it's *in principle* impossible to produce counter-examples to my stipulations. However, as long as counter-example haven't been actually contrived, there is no reason why I should operate on the assumption that their production is *possible*. I may, that is to say, tentatively assume that there exists no counter-example to my stipulations.

19. Summary

Let us now turn to summarize the discussion of negation of actions. Negation of an action consists in refraining from it.

The truth-table for negation is as follows:

$$\begin{aligned} \text{If } -.5 \leq v[A] \leq .5, v[\neg A] &= -1.7 v[A] \\ \text{If } v[A] > .5, v[\neg A] &= -1.7(1-v[A]) \\ \text{If } v[A] < -.5, v[\neg A] &= -1.7(1+v[A]) \end{aligned}$$

³⁵ Actually there is another difficulty for the objection at hand, only it doesn't undermine the latter. At any rate, this difficulty concerns the gradual entry into the moral realm. The objection implicitly assumes that gradually entering the moral realm is *inevitably* tied to an increase or decrease in the moral value of acts. But this isn't so. For actions may be borderline cases of morality or definitely within it and yet be morally neutral, because they enjoy equal degrees of rightness and wrongness.

Appendix: Reply to Professor Adler

1. Professor Adler's Worry

In conversation, Professor Jonathan Adler raised action-theoretic considerations against the view that an agent's foreseen sacrifice affects the moral value of his actions. Following is what I hope is a fair representation of Professor Adler's objection.

Our language encourages us to think that actions and events recur: we say things such as that Jack and Jill got married in May, and Dolly and George did the same thing in June.¹ The fact of recurrence in turn suggests that one and the same action (or event) may be performed (or occur) more than once. Now (the objection continues) suppose that *A* and *B* are—respectively—rich and poor, and that each contributes \$50 to famine relief. Then if an agent's foreseen sacrifice affects the moral value of his actions, *B*'s action enjoys a higher degree of rightness because his foreseen sacrifice is much more significant than *A*'s. This implies that *B* and *A* don't perform the same *moral* action, for how could the same moral action have different moral values? Yet this is counter-intuitive: we are inclined to think that *A* and *B* do perform the same moral action, as they perform the same physical action. We must accordingly conclude that an agent's foreseen sacrifice doesn't affect the moral value of his actions. More explicitly,

1. If an agent's foreseen sacrifice affects the moral value of his actions, then *A* and *B* don't perform the same moral action.
2. If *A* and *B* perform the same physical action, they perform the same moral action.
3. *A* and *B* perform the same physical action (as the fact of recurrence suggests).
4. From (2) and (3) it follows that *A* and *B* perform the same moral action.
5. From (1) and (4) it follows that an agent's foreseen sacrifice doesn't affect the moral value of his actions.

This doesn't mean, however, that there is no moral difference between *A* and *B*.

¹ The example is Donald Davidson's ("Events as Particulars").

Surely, *B* is more *praiseworthy* than *A*, since his action requires from him a significant sacrifice, whereas *A*'s action requires from *him* only a negligible sacrifice. Analogous remarks can be made with respect to wrong actions. If *A* and *B* both rob a bank, but (as before) *A* is rich and *B* is poor, then even though *A* and *B* perform the same act—the same *moral* act, that is—*A* is more *blameworthy* or less excusable than *B*. Or if *A* and *B* both refrain from contributing \$50 to famine relief, then even though the values of their respective omissions are the same, *B* is less blameworthy or more excusable than *A* because contributing \$50 requires from *B* a more significant sacrifice than it requires from *A*. So runs Professor Adler's objection.

2. Premise (3): Metaphysics (1)

Turn now to evaluate the objection. I shall begin with premise (3), which says that *A* and *B* perform the same physical action (as the fact of recurrence suggests). Well, in the first place one may follow Davidson in maintaining that recurrence doesn't necessarily require us to hold that the *same* action or event occurs more than once and hence doesn't require invoking events and actions as types or universals: recurrence requires holding only that *similar but distinct* events or actions (particular and unrepeatable) follow one after another.² Doing so would allow one to avoid the potentially embarrassing claim that one and the same action has more than a single moral value.

3. Premise (3): Metaphysics (2)

Secondly there is Professor Adler's reason for embracing premise (3). (If I'm not mistaken) it is that the causes and effects of an action don't play a role in its

² *Ibid.*, 184.

individuation. because they are not parts of it: if *A* and *B* are both strolling through the streets of Bologna, they are the same action, even if they are caused in very different manners and have very different effects. Likewise, if *A* and *B* are both contributing \$50 to famine relief, they are the same action, even if they are brought about (motivated) in different ways and have different consequences.

Not only does there appear to be no reason for thinking—as Professor Adler does—that actions and events may not be individuated by factors which are not parts of them, there’s an interesting metaphysical reason for thinking that they *may*. It is that since events have a unique position in the causal nexus, the latter provides for them a useful framework of identification and description, in very much the same way that the space-time coordinate system provides such a framework for material objects.³ (Note that since this statement refers to *all* actions—moral *and* non-moral alike—it in effect suggests that it should be relatively easy to turn a system from a calculus of only moral actions to a calculus of actions in general, where an action’s numerical value stands for its degree of rationality.)

4. Premise (3): Morality

In addition to that, Professor Adler’s view on event individuation has undesirable moral implications. Allowing that causes and effects of actions don’t play a role in their individuation amounts to excluding subjective (or foreseen) and objective (or actual) consequences of moral acts from playing a role in their individuation and thus in determining their moral value.⁴ (Indeed, this is precisely why Professor Adler insists that

³ See Davidson, “The Individuation of Events,” 179-80. In this essay Davidson advances the view that events are identical if and only if they have exactly the same causes and effects.

⁴ Foreseen consequences (or foreseen effects) are motives for acting and refraining from acts. Far from being effects of actions then, foreseen effects are causes of acts.

an agent's foreseen sacrifice doesn't affect the moral value of his actions. Foreseen sacrifice is after all part of the foreseen consequences of actions. If the latter are morally irrelevant, so is the former.) But if consequences (foreseen or actual) don't affect the moral value of actions, what does? What makes an action right or wrong? It must be some feature of acts other than their consequences, a feature such as being in accordance with the divine command. In other words, an anti-consequentialist stand on event individuation allows for no relativization whatsoever of moral value to consequences, thereby committing one to moral absolutism. (To *absolutism*, not merely to deontology. A deontological theory—such as W. D. Ross's—has it that certain actions are right or wrong in themselves regardless *to some extent* of the consequences of their performance or non-performance.⁵)

This result is undesirable on both a methodological and a philosophical ground. The methodological consideration is that absolutism is completely at odds with common moral and legal practices; in practice, consequences play a major—perhaps even exclusive—role in determining the moral value of actions. Since I am committed to being in line with common moral and legal practices, I cannot adopt absolutism or any view that entails it.

As for the philosophical ground, it includes several considerations, but I shall focus only on two of them. The first is that—as I argued in Chapter 1, Section 3—absolutism conflicts with the notion that rightness and wrongness admit of degrees. Now you will recall that the reason for the conflict is that according to both Kant's and religious absolutism, moral concepts are anchored in binary properties. One might therefore argue that it is in principle possible to offer a version of absolutism according to

⁵ *The Right and the Good*, 18.

which rightness rests on a *morally-relevant* property that admits of degrees. I agree. However, the burden of formulating such a position clearly lies with the absolutist.

As of now then, there exists no version of absolutism according to which rightness rests on morally relevant properties that admit of degrees. Consequentialism on the other hand makes room for scalar rightness and wrongness. This presumably means that rightness is scalar *only* because many of the properties to which consequences of actions ultimately reduce—such as pleasure and pain—admit of degrees. Or indeed, that only consequentialism can accommodate moral gradation.

Since absolutism excludes moral gradation, it entails that keeping a trivial promise and saving a life are equally right, that murder and theft are equally wrong, and so on. Note that this problem cannot be avoided by deferring moral gradation to praiseworthiness and blameworthiness. Claiming that *A* and *B* in the initial example differ not in the moral value of their actions but rather in their degree of praiseworthiness is at least *prima facie* plausible; but it is completely implausible to maintain that keeping a trivial promise and saving a life are of the same rightness degree, only their respective performers differ in their degree of praiseworthiness. In other words, moral gradation can perhaps be deferred to praiseworthiness and blameworthiness as long as the actions in question are (broadly and loosely speaking) of the same type (the comparison is made *within* a type). But once the comparison is made *between* types of action, moral gradation can no longer be captured in terms of blameworthiness and praiseworthiness.

The second philosophical consideration concerns moral dilemmas. Suppose that *A* and *B* are dilemma-generating actions, where refraining from *B* has much graver consequences than refraining from *A*, so that intuitively *B* is a more pressing duty than *A*.

Then if Professor Adler is correct and consequences of actions are morally irrelevant, there appears to be no way of *accounting* for this intuition.

5. Premise (2): Morality (1)

Interestingly, premise (2) of Professor Adler's objection (which states that if *A* and *B* perform the same physical action they perform the same moral action, and which—as I will shortly suggest—results in part from inappropriately placing metaphysics above ethics) also yields an implausible treatment of moral dilemmas. For suppose that *A* and *B* are dilemma-generating actions, and that both of them are breaking into a burning building in order to free a trapped person. Suppose further that while *A* requires a negligible sacrifice from the agent, *B* requires a considerable one (because—let us assume—*B* is much riskier to the agent, but in a way that doesn't affect the likelihood of her freeing the trapped person). If the agent's foreseen sacrifice doesn't affect the moral value of her actions, *A* and *B* enjoy the same moral value. This means that opting for *A* (and refraining from *B*) has the same moral value as opting for *B* (and refraining from *A*). But this is clearly not so: choosing *A* is more right than choosing *B*, because *A* is likely to preserve more lives. Since the only difference between *A* and *B* is the sacrifice that they require from the agent, the difference in their moral values must result from that difference. *Pace* Professor Adler then, an agent's foreseen sacrifice does affect the moral value of his actions.

6. Premise (2): Morality and Metaphysics (1)

Another problem for premise (2) surfaces when we look into the question of why think that since (as we have agreed to suppose) *A* and *B* perform the same physical action, they also perform the same moral action. Or (what is the same question) why

think that physical criteria of action individuation determine moral criteria? The view enjoys no *prima facie* plausibility. If anything, it's the opposite position (which says that moral criteria of action individuation are prior to—or determine—physical criteria) which appears to enjoy *prima facie* plausibility; for the need to handle issues such as moral responsibility is among the major motivations for individuating actions in the first place. This means that (the way things stand) one could well argue that since *A* and *B* (in the initial example) perform different moral actions, they also perform different physical actions.

7. Premise (2): Morality and Metaphysics (2)

A second way to challenge premise (2) then is to deny that physical individuation criteria determine moral criteria and not vice versa. A third way is to deny that physical individuation criteria *determine* moral criteria. For what intuition do we betray when we hold that even if *A* and *B* perform the same physical action (which is contributing \$50 to famine relief), they nonetheless perform different moral actions? Or when we hold that even if a four-year-old and Jackson Pollock both pour paint on a canvass thereby carrying out the same physical action, they nonetheless perform different acts from an aesthetic point of view? I can think of none. (Question: what on this view would count as *A* and *B* performing the same moral action? Answer: equivalent sacrifices for equivalent causes. Indeed, different physical actions. Thus, in order for Bill Gates to perform the same action that *B* (in the initial example) does, he must contribute a much more substantial amount of money.⁶) At the same time, I *can* think of intuitions that we betray when we subscribe to premise (2). Three such intuitions were already mentioned, and a fourth one

⁶ See Chapter 1, Section 8 for further discussion of this matter.

follows.

8. Premise (2): Morality (2)

Suppose that *A* and *B* both jump into a tempestuous river in order to save the life of a drowning person (so that they both perform the same action for the same cause). Suppose further that *A* is an excellent swimmer whereas *B* is a very poor swimmer. If Professor Adler is correct and the foreseen consequences of an action (including an agent's foreseen sacrifice) do not affect its moral value, then *A* and *B* perform actions of the same moral value. Intuitively however they do not: intuitively, *A*'s action is right while *B*'s is wrong (*B* is unlikely to save the drowning person and is likely to drown himself).

The proponent of the anti-consequentialist stand on action individuation will now retort that I have left out an important element of his position, and that is the claim that in such cases the moral difference between agents can be captured in terms of praiseworthiness and blameworthiness. In the present case (the proponent goes on) *A* and *B* perform the same moral action (and hence either both perform a right action or both perform a wrong one), but while *A* is praiseworthy *B* is blameworthy.

This reply will obviously not do, for it implies either that *A* is held praiseworthy for a wrong act, or that *B* is held blameworthy for a right one. (As Professor Michael Levin pointed out to me, holding a person praiseworthy for an essentially wrong or irrational action is a somewhat common practice: we *do* sometimes deem others brave—that is, praiseworthy—for actions (such as a Bungee jump) that involve a considerable risk for an unworthy cause; which means that there *is* a point in keeping the categories 'right' and 'praiseworthy' separate. Well, even if this is a reason for keeping 'right' and

‘praiseworthy’ separate—and I think that it isn’t, because deeming actions such as a Bungee jump praiseworthy is an anathema to me—it’s by no means a reason for burdening the ill-equipped ‘praiseworthy’ with capturing moral gradation.)

9. General Difficulties

In addition to the problems afflicting premises (2) and (3) of Professor Adler’s argument, there are difficulties for more general aspects of his position. One is that it implies that an agent’s degree of blameworthiness or praiseworthiness is irrelevant to the moral value of her actions (recall that according to Professor Adler, *A* and *B* in the initial example perform actions of the same moral value yet differ in their degree of praiseworthiness). This result is not only odd in and of itself, it also flies in the face of common practice (to which I am committed). For in practice, an agent’s blameworthiness *does* affect the moral value of her actions: if an agent commits a homicide for which she may hardly be held blameworthy (because it was done in defense of herself or her family, for example), then her act is considered a justifiable homicide, which—judging by the punishment that the law provides for it—is an offense of a lower wrongness degree than a homicide for which the agent may be held highly blameworthy (felonious homicide).⁷

Another (and here last) difficulty is that if an agent’s foreseen sacrifice doesn’t affect the moral value of his actions, then saving a life at risk of *my* life is of the same moral value as saving that life at risk of *your* life. Since at least on some occasions the

⁷ A homicide is justifiable if committed with full intent, but under such circumstances as to render it proper and necessary, as where the proper officer executes a criminal in strict conformity with his sentence, or where one kills another in defense of himself or his family. A homicide is felonious when committed willfully and without sufficient justification. This includes suicide, manslaughter, and murder (Erik Beckman, *The Criminal Justice Dictionary*, 101-2). Since both kinds of homicide are the killing of a human being for one’s own advantage, they are instances of the same type of action.

former would be right and the latter wrong (i.e., the actions would be of *different* moral values). we must conclude that an agent's foreseen sacrifice does affect the moral value of his actions.

10. Professor Adler's Reply

I bring Professor Adler's response verbatim:

...I thought you are claiming that "foreseen sacrifice" *always* affects the moral value of the act. So I thought that I only need to establish that they are sometimes irrelevant. So I do not have to hold to premise (2), nor premise (1) for each pair *A* and *B* (i.e. as variables). In short, I think that there are *some* cases where *A* and *B* do perform the same moral action, such as lying when asked if they had eaten the last cookie. They both say "no." But *B* is more resistant to answering "no" because he tells the truth more habitually than *A*. Now I think there is a way of filling in the background so that they will have performed the same moral action and they have both acted (equally) wrongly. Nevertheless, if *B* lied after more of a struggle than *A*, then we might praise *B* more than *A* (or condemn *A* more than *B*). In any case, what is at issue is whether you claim that *necessarily* if *A* and *B* do differ in ways that bear on their degree of resistance to a moral (immoral) act that it follows that they have not performed the same act.⁸

11. Reply to the Reply

The view that foreseen sacrifice *never* affects an act's moral value—the view which I took Professor Adler to be advancing—has an interesting consideration in its favor: namely, that actions and events may not be individuated by factors that are not parts of them. Since this consideration applies to all consequences, it cannot be invoked in support of the position—encapsulated in Professor Adler's reply—that foreseen sacrifice doesn't affect moral value only on *some* occasions. It appears therefore that

⁸ In fairness to both Professor Adler and myself I should add that the response concludes as follows. "Even if my interpretation of you is correct and so my objection can be formulated in much weaker terms, there is a great deal in your note that survives."

unless there is another consideration in its support, this position has nothing to recommend it.

And, yes. I do think that foreseen sacrifice *always* affects the moral value of actions. (Indeed then, a single counter-example would suffice for refuting my position. Professor Adler hasn't produced one in his response, nor does he claim to have done so: he only gestures at the direction in which one might proceed: the background remains to be filled in.) As I argue in some detail in Chapter 1, Section 8, I think that the sacrifice involved in performing an action is crucial for distinguishing rightness from goodness and wrongness from badness.

Chapter 3: 'Or'

1. Preliminary Remarks (1)

Morality (I take it) is concerned first and foremost with guiding actions: with providing answers to the question 'What should I do?' Now this question frequently arises in connection with non-moral activities. A chess player often wonders what she should do to achieve a checkmate, a decorator—to gain a certain aesthetic effect (harmony perhaps), and an academic—to get her work published. Which means that the opening statements must be qualified. Morality is concerned with answering not simply the question 'What should I do?' but rather the question 'What should I do in order to realize a certain *moral* value?'

True enough, in recent years (since the early 1980s, I believe) it has been suggested more than once that the subject matter of morality (or rather ethics¹) is organized in the first instance around the question 'How should I *live*?' Just as Socrates (Plato?) proposes.² Proponents of this position argue that it makes room within ethics for realization of not only moral but also *non-moral* values such as an intelligent understanding of the world. They argue in other words that the position makes room for permission to sometimes refrain from realizing moral values, and pursue instead projects of the kind that tends to give at least some persons a reason for living at all.³

If I'm fortunate enough to live in the way that *I* in my opinion should live, I am

¹ See Bernard Williams's *Ethics and the Limits of Morality*.

² *Republic*, 352D.

³ If 'moral value' means satisfaction of rational *personal* desires—as it does in ethical egoism—then moral value largely overlaps or even converges (which of the two depends on what one means by 'rational') with values that give persons a reason for living.

Despite its important affinity to egoism, the view that persons may sometimes pursue their own projects is regularly associated primarily with deontology (see James Dreier's "Structures of Normative Theories" for a discussion of a closely related matter). A fact which I find curious.

likely to examine whether the position in question rests on a plausible view of the relationship between ethics and the meaning of life (I suspect it doesn't). Until that time comes, I shall have to simply *assume* that the main focus of ethics is the question 'What should I do to realize this or that *moral* value?'

Of the kinds of issues that invite an answer to this question, moral dilemmas take center stage in both philosophy and belles lettres (in life too, though perhaps to a lesser extent). Examples abound. In Plato's *Apology*, Socrates is depicted as arguing that if the state spares his life provided that he ceases teaching, he would be torn between—on the one hand—his duty to obey the state, and—on the other hand—his duty (assigned to him by the god Apollo) to carry on his teaching which he deems necessary for the true good of the state. Sophocles's *Antigone* is faced with a choice between the religious duty to give her brother Polyneices proper burial and the civil duty to refrain from that act.

Sartre tells of a student, who during the Second World War

...sought me out in the following circumstances. His father was quarrelling with his mother and was also inclined to be a "collaborator"; his elder brother had been killed in the German offensive of 1940 and this young man, with a sentiment somewhat primitive, but generous, burned to avenge him. His mother was living alone with him, deeply affected by the semi-treason of his father and by the death of her eldest son, and her one consolation was in this young man. But he at this moment had the choice between going to England to join the Free French Forces or of staying near his mother and helping her to live. He fully realized this woman was living only for him and that his disappearance—or perhaps his death—would plunge her into despair. He also realized that, concretely and in fact every action he performed on his mother's behalf would be sure of effect in the sense of his aiding her to live, where as anything he did in order to go and fight would be ambiguous action which might vanish like water into sand and serve no purpose.⁴

⁴ "Existentialism and Humanism," 400.

The examples demonstrate that moral dilemmas arise because the agent's moral beliefs imply that she ought to do both *A* and *B* where the circumstances are such that doing *A* necessitates refraining from *B*, and vice versa.⁵ Dilemmas thus consist in a pair of conflicting actions or courses of action, both of which are called *for*.

Now since wrong actions are by no means called *for* but are instead "called against" (regardless of whether they are also *forbidden*), pairs of wrong actions never generate dilemmas. But one might complain that if the sheriff of a small town can prevent hundreds of deaths by executing an innocent man as a scapegoat, then letting hundreds be slaughtered (*A*) and executing an innocent (*B*)—both of which wrong actions *par excellence*—are both called *for*.

I answer that under the circumstances just described, killing an innocent man just *is* preventing hundreds of deaths. Far from showing therefore that *pairs* of wrong actions sometimes generate dilemmas, the scenario under consideration in fact demonstrates that on certain occasions there are reasons for both performing and refraining from *one and the same* act. (Since this is impossible, such occasions no doubt *appear* dilemmatic. Not only that, they may in fact *be* dilemmatic, a thought I explore in Chapter 5, Section 12).

Another way to meet the objection at hand is to point out that *A* and *B* amount (respectively) to refraining from executing an innocent man and preventing hundreds of deaths, both of which are *right* actions. The reasons why I much prefer the first reply emerge in Chapter 5.

⁵ Choice situations that involve more than two options are considered in Section 10 below.

2. Preliminary Remarks (2)

Dilemmas then consist in pairs of conflicting actions each of which is required. Presented with a moral dilemma, an agent must choose between *either* the one *or* the other horn (she cannot perform—and shouldn't refrain from—both). While this remark may be superfluous content wise, as far as form is concerned it's quite telling. It reveals that in moral discourse 'or' serves to express dilemmas. This may or may not be the *only* role played by 'or' in ethical contexts: it is however the only role on which I shall *focus*.⁶ Let's call a dilemma-expressing 'or' 'dilemmatic.'

While negation maps actions into (onto?) their negations, dilemmatic disjunction maps pairs of actions into dilemmatic-disjunctive actions. Negation of an action amounts to refraining from it. What does dilemmatic-disjunctive action amount to? To a first approximation, it amounts to choosing or deciding among a pair of conflicting actions, both of which are required. 'To a first approximation' because—since choosing is in and of itself a *mental* act—worries of the metaphysical as well as the ethical variety arise.

As far as metaphysics is concerned, if the range of dilemmatic disjunction consists in actions, how could they be mental—as opposed to spatio-temporal—entities? I clearly must forgo exhaustive pursuit of this question here. Briefly however, if the mental is ultimately physical, then mental acts *are* spatio-temporal entities, only they take place "inside the body." Besides, few deny that mental *events*—such as perceiving and remembering—occur. Since actions are presumably nothing but events of a sort, mental acts ought to be admitted as denizens of the world. Moreover, 'mental acts' seems to best describe (among others) decision-making and problem-solving. And finally, since

⁶ More on this claim in Section 13 below.

decisions and refrainings are ontologically on a par (refraining after all doesn't occur "outside the body"⁷), if one is unworried by letting the latter inhabit the range of negation, one should be equally content with allowing the former constitute the range of dilemmatic disjunction. The metaphysical worry then doesn't call for adjustments in the suggestion that dilemmatic-disjunctive action amounts to choosing or deciding.

The ethical complaint does, however. If Sartre's student opts for joining the Free French Forces but doesn't act upon his decision, then for all intents and purposes he has chosen to stay near his mother. To be a choice-to-join-the-Forces, the student's choice must be accompanied by at least a sincere attempt to realize it (some would say: a *successful* attempt).⁸ Dilemmatic-disjunctive action amounts accordingly not simply to choosing among a pair of conflicting duties, but to choosing followed by a sincere attempt to act upon the choice.⁹

To avoid cumbersome wordings however, I shall often mention either the choice or the attempt alone. Moreover, since I assume that objective consequences are morally irrelevant, I in effect hold that successful and unsuccessful attempts to perform an act are morally on a par. (Indeed, as noted in Chapter 1, Section 7, my assumption concerning objective consequences entails that murder and attempted murder are equally wrong.) For this reason I shall usually replace 'performing *A*' for 'attempting to perform *A*.' Thus, I shall use 'choosing and attempting to perform *A*,' 'choosing *A*,' 'attempting to

⁷ Philippa Foot's work requires a qualification: *some* instances of refraining don't occur in space.

⁸ This account bears affinity to Jennifer Hornsby's views (*Actions*). Hornsby argues that causally basic actions are the "tryings" which are the inner causes of bodily movements. However, tryings that fail to issue in bodily movements are not actions.

⁹ A further qualification will be introduced in Section 13.

perform A and 'performing A ' interchangeably.

3. Opportunity Cost

The evaluation function for disjunction in both two-valued logic and in popular—that is, Łukasiewicz's¹⁰—many-valued logics is:

$$v[p \vee q] = \max \{v[p], v[q]\}$$

($v[p]$ is p 's degree of truth which we suppose is a real number between 0 and 1). In words, the degree of truth of a disjunction is the maximum of the degrees of truth of its disjuncts. In these systems then, there is no need to take into consideration negation values of disjuncts in order to determine the truth-value of disjunction.

Dilemmatic disjunction by contrast *should*—in one way or another—be sensitive to the truth-table for negation. To see why consider Sonja. She is faced with a choice between two required yet mutually exclusive actions, A and B . A 's degree of rightness is .5, B 's is .8. Because B enjoys the higher rightness degree, Sonja might at first believe that *it* is the act she should opt for. However, since B excludes A , its performance yields not only .8 rightness degrees but also .85 degrees of *wrongness* ($v[\neg A] = -.85$). On the other hand, doing A (and avoiding B) would produce .5 degrees of rightness and .34 wrongness degrees (-.34 being the value of refraining from B). Thus, whereas the overall value of performing B is negative, that of performing A is positive. Which presumably means that under the circumstances in question, Sonja ought to opt for A , not for B . Contrary to what at first blush seemed to be morally required.

That she should indeed select A becomes instantly clear once concrete actions

¹⁰ See Jan Łukasiewicz and Alfred Tarski, "Investigations into the Sentential Calculus." This family of systems is often taken to provide the semantics for the logical connectives in degree theories. See for example Timothy Williamson, *Vagueness*, 115-8 (Williamson presents this semantics in order to critique it). Or Greame Forbes, "Thisness and Vagueness."

substitute for the variables. Let A be breaking—at a negligible cost to the agent—the door lock of a burning building in order to free a trapped person. Let B be jumping into a tempestuous river to save a drowning person (Sonja is an excellent swimmer, or else her action would be wrong regardless of whether A is also called for and available to her). Then it's immediately evident that Sonja ought to execute the action that is likely to preserve more lives. And that is A .

What if A and B were (respectively) of value .3 and .35? $v[\neg A] = -.51$ while $v[\neg B] = -.595$. Since doing B would produce less overall wrongness, *it* is the act Sonja should pick. This time round then the horn she should choose is also the act which enjoys the higher rightness degree.

A thorough treatment of moral dilemmas then requires taking into account not merely the rightness degree, but also the negation value, of each horn. This conclusion echoes a tenet fundamental to economics: namely, that analyses of decision-making processes should be made with an eye to alternative opportunities. Specifically, given the scarcity of time and other resources, nearly all decisions or choices—whether of individuals or organizations—involve trade-offs. A consumer with a fixed income for instance may have to forgo a vacation or several suits of clothes in order to purchase a new home appliance. The alternative opportunities forgone in the choice of one good or activity over others is the opportunity cost of that choice. Drawing on this terminology one could say that, in handling moral dilemmas, agents should take into consideration the expected opportunity cost of its possible solutions.

4. Fixing Ideas (1)

I have indicated *why* dilemmatic disjunction should be sensitive to the truth-table for negation. I have not yet considered the *how*. Previous remarks concerning this problem are admittedly somewhat tendentious. Letting \mathbf{V} signify dilemmatic disjunction, and letting $A \mathbf{V} B$ read ‘opting for A given that B is also called for and available to the agent,’ the foregoing discussion suggests that the value of $A \mathbf{V} B$ is $v[A] + v[\neg B]$, while the value of $B \mathbf{V} A$ is $v[B] + v[\neg A]$. More formally,

$$(i) \quad v[A \mathbf{V} B] = v[A] + v[\neg B]$$

$$(ii) \quad v[B \mathbf{V} A] = v[B] + v[\neg A]$$

This offer is certainly attractive. Suppose that A —an act of value .6—is saving a life at a rather negligible cost, whereas B —an act enjoying .1 rightness degrees—is preventing a minor harm at a negligible sacrifice. And recall that A ’s negation value is -.68 while B ’s is -.17. Then the truth-table at hand outputs .43 and -.58 for $A \mathbf{V} B$ and $B \mathbf{V} A$, respectively. Which is to say that preferring B over A (A over B) is—as it should be—wrong (right).

5. Two-valued Utilitarianism

Very worth noting is the insight that this case and its like give into the rationale behind two-valued utilitarianism. You will recall that according to this theory, agents ought to not merely produce—but rather *maximize* (optimize)—utility: they are required to carry out that act which, of all the alternatives open to them, would produce not simply a good—but rather the *best*—outcome. Since this view entails that failing to bring about the best results is wrong even if one’s act in fact generates *beneficial* consequences,

Marcus Singer for one has found it objectionable.¹¹ Surely however, opting for *B* given *A* is wrong *even though* it promotes *goodness*. (Yet another reason for denying that rightness is identical with goodness.) Under certain circumstances therefore, agents should indeed aim at producing no less than *maximum* (or optimum) utility.

Not that I subscribe to this particular theory. I think that when faced with a moral dilemma, agents should seek to maximize not utility, but moral value instead. (In addition I deny that all failures to do so are wrong. Explanation is forthcoming.) Here is why. If a man can, at equal risk to himself, save either his son or two strangers, he may or even should prefer his child over the two. While such a position appears to collide with two-valued utilitarianism, it sits well with a theory that requires maximization of moral value. Here is how. Because people normally love their offspring, the sacrifice involved in saving one's son (*A*) is smaller than that involved in saving two strangers (*B*). Let's assume that *A* and *B* are carried out for an equally worthy cause (allow that saving one's own son is as worthy a cause as saving two strangers). Then if $v[A] = .8$ and $v[B] = .9$, the man should choose *A* ($.8 - .17 = .63$, whereas $.9 - .34 = .56$).¹² However, opting for *B* is by no means wrong: it involves after all refraining from saving one's own flesh and blood in order to save two lives. This is why I deny that all failures to maximize moral value are wrong.

6. Fixing Ideas (2)

Returning now to the valuation procedure under consideration, I have established

¹¹ "Actual Consequence Utilitarianism."

¹² In case there are utilitarian grounds for thinking that *A* and *B* are morally equivalent (or that *A*'s moral value exceeds that of *B*), my view concerning moral dilemmas overlaps with two-valued utilitarianism to an even greater extent. However, its merits notwithstanding, two-valued utilitarianism by itself is an inadequate moral theory, as it is incapable of capturing moral gradation.

that it yields desirable results for some pairs of actions. Trouble is: for other pairs it yields *undesirable* output. Consider *A* and *B*, where *A*—of value .2—is keeping a trivial promise at a negligible price and *B*—of .3 degrees of rightness—is fulfilling a rather important duty at a negligible cost. $v[A \vee B] = .2 - .51 = -.31$ while $v[B \vee A] = .3 - .34 = -.04$. Which means that choosing and attempting to carry out the lesser of two evils— $B \vee A$ —is wrong (albeit *slightly* wrong). Or that one acts wrongly regardless of the avenue he pursues.

Similar discontent arises upon plugging .5 and .6 into the table ($.5 - .68 = -.18$ and $.6 - .85 = -.25$). And .5 and .5 ($.5 - .85 = -.35$). The latter case and its like are especially frustrating, because here it matters not which horn the agent decides on: all that matters is that she doesn't refrain from both: as long as she carries out one of them, she acts appropriately.

Is it the case then that the valuation procedure produces unsatisfactory output only for value pairs whose members are separated by an insubstantial amount of rightness? Not at all. Plugging into the table .3 and .9—a value pair whose members are separated by a *substantial* amount of rightness—entails that one acts rightly regardless of the horn she opts for ($.3 - .17 = .13$, whereas $.9 - .51 = .39$). One might feel that the present is quite unproblematic a result, because an act of value .9 involves a significant sacrifice. Well, not quite. While refraining from an act that involves a significant sacrifice is only *mildly* wrong, it is nonetheless wrong. And it remains so, even if the refraining enables the agent to produce a crumb of goodness. A lifeguard on duty who—in place of risking his life to save a life—prevents a minor harm, without a doubt acts wrongly.

So far I have enumerated problems that arise for only *some* value pairs. A difficulty for *all* pairs is the fact that choosing the required act (maximizing moral value) is never as right as that act itself (suppose A ought to be preferred over B . Then $v[A \vee B] = v[A] + v[\neg B]$. Since $v[\neg B]$ is without exception negative, $v[A \vee B]$ is always lower than $v[A]$). Put differently, agents are in a sense punished for appropriately resolving dilemmas which (I am assuming) they involuntarily encounter.

In sum, the valuation procedure under consideration fails to meet four fundamental conditions. The first is that all instances of rightness maximization (or wrongness minimization) be right.¹³ The second—that choosing between morally equivalent acts be right.¹⁴ The third is that failing to maximize rightness be wrong,¹⁵ unless the act one performs involves a material sacrifice and a worthy cause. Specifically, if the act the agent should give up is highly right, opting for it should also be right (though not to an unusual degree).¹⁶ Finally, opting for the required act (maximizing net moral value) ought to be at least as right as that act itself.¹⁷

7. Evaluation Procedure

I proceed to offer an evaluation function which answers all four conditions, as well as a fifth one: namely, that the value assigned to refraining from both horns (in symbols: $v[\vee A, B]$) equal the value of refraining from A and refraining from B (in symbols: $v[\neg A \wedge \neg B]$).

If $v[A] + v[\neg B] > v[B] + v[\neg A]$, then

¹³ Refer to the first three value pairs in this Section.

¹⁴ Refer to the third value pair.

¹⁵ Refer to the fourth value pair.

¹⁶ The proviso concerning highly right actions was motivated at the end of Section 5. Additional motivation is provided below.

¹⁷ Refer to all value pairs.

- (i) $v[A \vee B] = v[A]$
- (ii) $v[B \vee A] = v[B] + v[\neg A]$

If $v[A] + v[\neg B] < v[B] + v[\neg A]$, then

- (i) $v[A \vee B] = v[A] + v[\neg B]$
- (ii) $v[B \vee A] = v[B]$

If $v[A] + v[\neg B] = v[B] + v[\neg A]$, then

- (i) $v[A \vee B] = v[A]$
- (ii) $v[B \vee A] = v[A]$

If $v[\neg A] > v[\neg B]$ or $v[\neg B] > v[\neg A]$, then

- (i) $v[\vee A, B] = v[\neg A \wedge \neg B] = \frac{1}{2} (v[\neg A] + v[\neg B])$ ¹⁸

Let's run a few adequacy tests. If $v[A] = .9$ and $v[B] = .6$, then—since the agent should perform B ($.6 - .17 = .43$, whereas $.9 - .68 = .22$)— $v[A \vee B] = v[A] + v[\neg B] = .22$ and $v[B \vee A] = v[B] = .6$. Which means that deciding on B is as right as B itself, while choosing A —an act involving a substantial sacrifice—is slightly right. One might object that choosing A given that B is also available is not even slightly right, but in fact wrong. After all, risking one's life to save a life is wrong if the agent can instead save a life at little or no risk to himself. Well, in the first place, by assigning to $A \vee B$ only a *slight* degree of rightness, the function at hand ultimately respects to some extent the intuition motivating this objection. Secondly, though I can certainly feel the force of the objection, arguments to the contrary have more appeal in my eyes. One such argument appears at the end of Section 5. Another is that acts which involve much sacrifice for a worthy cause are heroic (whether or not they are also obligatory) and hence right: it

¹⁸ Explanation and discussion of this line must await until Chapter 4, Sections 9-10 and 18.

stands to reason that they should remain right even where their performance is irrational.¹⁹

Next. If $v[A] = v[B] = .6$, then $[A] + v[\neg B] = v[B] + v[\neg A] = v[A] = .6$. Meaning that choosing between morally equivalent acts is right, and that opting for the required act is at least as right as that act itself.

If $v[A] = .2$ and $v[B] = .3$, then—since the agent should choose B ($.3 - .34 = -.04$, whereas $.2 - .51 = -.31$)— $v[A \vee B] = v[A] + v[\neg B] = -.31$ and $v[B \vee A] = v[B] = .3$. That is, failing to minimize wrongness is wrong—an especially welcome consequence in light of the fact that B generates *much* less wrongness than A . In light of that same fact however, the low wrongness degree of choosing A ($-.04$) is admittedly somewhat unsettling. Yet another bullet I shall bite.

8. $A \wedge \neg B$

Since an agent who performs $A \vee B$ in effect commits A and refrains from B , it stands to reason that $A \vee B$ is equivalent to $A \wedge \neg B$. Likewise, since carrying out $B \vee A$ amounts to choosing B and refraining from A , $B \vee A$ is presumably equivalent to $B \wedge \neg A$. Is this indeed the case? The answer must of course await until we have given semantics to ‘and.’ See Chapter 4, Section 18.

9. Conditional Value and Commutation²⁰

Dilemmatic disjunction is conditional: its value changes depending on the action that agents choose and attempt to perform. In other words, unless A and B are equivalent,

¹⁹ But acts that involve much sacrifice for an *unworthy* cause *are* wrong.

²⁰ Throughout this work I understand ‘association’ and ‘commutation’ semantically. Thus, \vee commutes iff $v[A \vee B] = v[B \vee A]$. Likewise, it associates iff $v[A \vee (B \vee C)] = v[(A \vee B) \vee C]$.

opting for A given that B is also called for and available to the agent differs in value from choosing B given A . And in symbols, $v[A \vee B] \neq v[B \vee A]$, unless $v[A] = v[B]$.

Formally this means that dilemmatic disjunction fails to commute. In this respect it's unlike disjunction in either two-valued logic or popular many-valued logics. You will recall that in these frameworks $v[p \vee q] = \max \{v[p], v[q]\}$; which means that $v[p \vee q] = v[q \vee p]$.

10. n -lemmas

Dilemmatic disjunction (as both its name and the foregoing discussion suggest) is defined for moral *dilemmas*. How can its scope be extended so as to include choice situations consisting in just any number of options (n -lemmas)?

We need an example before our eyes. Suppose that Sonja is faced with a moral trilemma: she must choose between A , B , and C , where $v[A] = .5$, $v[B] = .6$, and $v[C] = .7$. As before, the agent should perform that act which, of all the alternatives open to her, would maximize net moral value. To identify the required act, Sonja needs to compute and compare three values. Namely:

- (1) $v[A] + v[\neg B] + v[\neg C] = .5 - .68 - .51 = -.69$
- (2) $v[B] + v[\neg A] + v[\neg C] = .6 - .85 - .51 = -.76$
- (3) $v[C] + v[\neg B] + v[\neg A] = .7 - .68 - .85 = -.83$

Since performing A would maximize net moral value (via minimizing wrongness), A is the act Sonja should pick. Previous practice dictates that opting for A shall enjoy A 's moral value or .5. degrees of rightness.

It may seem that adherence to previous practice dictates also that the value of opting for B (or C) equals $v[B] + v[\neg A] + v[\neg C]$ (or $v[C] + v[\neg B] + v[\neg A]$). But this need not—nor can it—be the case. As to why it need not, the answer is forthcoming.

Concerning why it cannot, there are a couple of reasons. In the first place, evaluation procedures that involve addition require an open ended interval (if this isn't obvious, think of a decalemma); but such an interval would undermine some of the most important stipulations made so far (I elaborate on this claim in Chapter 4, Section 14). Secondly, the procedure in question makes little philosophical and ethical sense, for it entails that the value of disjunctive actions decreases as the number of options increases. Since the number of options facing an agent is a morally irrelevant feature, yoking the moral value of her choice to it is intolerably arbitrary.

I can think of two ways to overcome these obstacles, both of which involve looking upon n -lemmas as *dilemmas* of a sort. The first (Proposal 1) states that opting for B (C) amounts to preferring B (C) over a single other action: namely, the product of A and C (A and B): the moral value of refraining from the product of A and C (A and B) is the average—indeed a product—of the negation values of these actions. Accordingly, the moral value of opting for B is $v[B] + \frac{1}{2}(v[\neg A] + v[\neg C])$, whereas the moral value of opting for C is $v[C] + \frac{1}{2}(v[\neg A] + v[\neg B])$. Note that since the average of $v[\neg A]$ is $v[\neg A]$, the present evaluation function is but a generalization of the function I introduced in Section 7 (you will recall that according to the latter, if $v[A] + v[\neg B] > v[B] + v[\neg A]$, then $v[B \vee A] = v[B] + v[\neg A]$). Therefore, previous practice dictates that the value of opting for B (or C) equals $v[B] + v[\neg A] + v[\neg C]$ (or $v[C] + v[\neg B] + v[\neg A]$) to the same degree it dictates that this value equals $v[B] + \frac{1}{2}(v[\neg A] + v[\neg C])$ (or $v[C] + \frac{1}{2}(v[\neg A] + v[\neg B])$).

According to Proposal 1 then, the evaluation procedure for a trilemma is:

if $v[A] + v[\neg B] + v[\neg C] > v[B] + v[\neg A] + v[\neg C]$ and $v[A] + v[\neg B] + v[\neg C] > v[C] + v[\neg B] + v[\neg A]$, then

$$(i) \quad v[A \vee (B, C)] = v[A]$$

- (ii) $v[B \vee (A, C)] = v[B] + \frac{1}{2} (v[\neg A] + v[\neg C])$
- (iii) $v[C \vee (B, A)] = v[C] + \frac{1}{2} (v[\neg A] + v[\neg B])$

if $v[B] + v[\neg A] + v[\neg C] > v[A] + v[\neg B] + v[\neg C]$ and $v[B] + v[\neg A] + v[\neg C] > v[C] + v[\neg B] + v[\neg A]$, then

- (i) $v[B \vee (A, C)] = v[B]$
- (ii) $v[A \vee (B, C)] = v[A] + \frac{1}{2} (v[\neg A] + v[\neg C])$
- (iii) $v[C \vee (B, A)] = v[C] + \frac{1}{2} (v[\neg A] + v[\neg B])$

if $v[C] + v[\neg B] + v[\neg A] > v[A] + v[\neg B] + v[\neg C]$ and $v[C] + v[\neg B] + v[\neg A] > v[B] + v[\neg A] + v[\neg C]$, then

- (i) $v[C \vee (A, B)] = v[C]$
- (ii) $v[B \vee (A, C)] = v[B] + \frac{1}{2} (v[\neg A] + v[\neg C])$
- (iii) $v[A \vee (B, C)] = v[A] + \frac{1}{2} (v[\neg B] + v[\neg C])$

if $v[A] + v[\neg B] + v[\neg C] = v[B] + v[\neg A] + v[\neg C] = v[C] + v[\neg B] + v[\neg A]$, then

- (i) $v[A \vee (B, C)] = v[A]$
- (ii) $v[B \vee (A, C)] = v[A]$
- (iii) $v[C \vee (B, A)] = v[A]$

When formulated generally, Proposal 1 states that if an agent picks the required (i.e., rightness maximizing) act, the value of her choice equals the value of that act; if she doesn't, the value of her choice equals the value of the act she opted for plus the average of the negation values of the remaining options.

While certainly coping with old problems (I refer of course to the pair of obstacles I mentioned earlier), Proposal 1 appears to generate a new difficulty. Suppose Sonja must choose between A and B , where $v[A] = .5$ and $v[B] = .6$. Since $.5 - .68 = -.18$ while $.6 - .85 = -.25$, she should opt for A . In line with the evaluation procedure I introduced in Section 7, $v[A \vee B] = v[A] = .5$ while $v[B \vee A] = v[B] + v[\neg A] = -.25$. Next suppose that Sonja should choose between A , B , and C , where $v[C] = .7$. Since performing A would

maximize net moral value (evidence to that effect appears at the outset of this Section). Proposal 1 yields that $v[A \vee (B, C)] = v[A] = .5$, $v[B \vee (A, C)] = v[B] + \frac{1}{2} (v[\neg A] + v[\neg C]) = -.08$, and $v[C \vee (B, A)] = v[C] + \frac{1}{2} (v[\neg A] + v[\neg B]) = .7 + \frac{1}{2} (-.85 - .68) = -.065$. It follows that preferring B over A and C isn't as wrong as preferring B over A alone. Yet it appears that preferring B is equally wrong in either case. The reason is that $v[\neg C]$ should not be taken into consideration at all, as refraining from C is desirable. In slightly other words, taking into account the negation value of a certain act suggests that refraining from it is *wrong*; yet given A and B , refraining from C is *right*. In short, as far as B is concerned, only A 's availability matters; C 's is irrelevant.

In response it could be argued that additional options complicate the process of decision making; hence failing to choose the required act given three options *should* be more excusable or less wrong than failing to do so given merely two possibilities. Ergo, C 's availability *does* after all make a difference.

Its allure notwithstanding, this response fails to sit well with Proposal 1. Let $v[D]$ be .4. Then

- (1) $v[A] + v[\neg B] + v[\neg C] + v[\neg D] = .5 - .68 - .51 - .68 = -1.37$
- (2) $v[B] + v[\neg A] + v[\neg C] + v[\neg D] = .6 - .85 - .51 - .68 = -1.44$
- (3) $v[C] + v[\neg A] + v[\neg B] + v[\neg D] = .7 - .85 - .68 - .68 = -1.51$
- (4) $v[D] + v[\neg A] + v[\neg B] + v[\neg C] = .4 - .85 - .51 - .68 = -1.64$

Which means that (again) A is the act agents should pick. If the suggestion currently under scrutiny were consistent with Proposal 1, opting for B given A , C and D would not be as wrong as opting for B given A and C . Yet $v[B \vee (A, C, D)] = v[B] + \frac{1}{3} (v[\neg A] + v[\neg C] + v[\neg D]) = .6 - .68 = -.08$ and $v[B \vee (A, C)] = v[B] + \frac{1}{2} (v[\neg A] + v[\neg C]) = .6 - .68 = -.08$.

Another possible response to the difficulty at hand is that preferring B over A and C is not as wrong as preferring B over A alone, because only the latter choice amounts to maximizing wrongness. (In other words, only preferring B over A and C involves a right action: namely, preferring B over C .) Well, if *this* reply cohered well with Proposal 1, then preferring C over A and B (which—as shown by the calculations at the outset—maximizes wrongness) would be as wrong as preferring C over A alone (between A and C , agents should opt for A , because $.5 - .51 = .01$ whereas $.7 - .85 = -.15$). But $v[C \vee A] = .5 - .51 = -.01$ while $v[C \vee (B, A)] = -.065$.

Proposal 1 then is faulty because it takes into consideration irrelevant moral values. Proposal 2, to which I now turn, is designed to amend for this shortcoming. Like its predecessor, it states that opting for the required act enjoys that act's degree of rightness. Unlike its predecessor, it stipulates that opting for the second (third, fourth, and so on) best option is judged with an eye to the required option alone. (This explains why Proposal 2 involves viewing n -lemmas as *dilemmas* of a kind.) For instance, if—as before— $v[A] = .5$, $v[B] = .6$, and $v[C] = .7$, then

- (i) $v[A \vee (B, C)] = v[A]$
- (ii) $v[B \vee (A, C)] = v[B] + v[\neg A]$
- (iii) $v[C \vee (B, A)] = v[C] + v[\neg A]$

One might object that Proposal 2 too is defective, not because it includes irrelevant values, but because it excludes relevant ones. Recall (the objection proceeds) that $v[A] + v[\neg B] + v[\neg C] = -.69$, $v[B] + v[\neg A] + v[\neg C] = -.76$, and $v[C] + v[\neg B] + v[\neg A] = -.83$. This presumably means that relative to C , B too—and not only A —is required. In assessing $v[C \vee (B, A)]$ therefore, $v[\neg B]$ too—and not only $v[\neg A]$ —should be taken into account. In particular, $v[C \vee (B, A)]$ should equal not $v[C] + v[\neg A]$, but

rather $v[C] + \frac{1}{2} (v[\neg A] + v[\neg B])$. An adequate treatment of n -lemmas then calls for a blend of Proposals 1 and 2 (Proposal 1½). Proposal 1½ states that opting for an act enjoys that act's moral value plus the average of the negation values of all and only relevant options. Thus, if $v[A] = .5$, $v[B] = .6$, $[C] = .7$, and $v[D] = .4$, then no act is relevant to A , only A is relevant to B , both A and B are relevant to C , and all acts are relevant to D . (Evidence to that effect appears a few paragraphs back.) Accordingly,

- (i) $v[A \vee (B, C, D)] = v[A]$
- (ii) $v[B \vee (A, C, D)] = v[B] + v[\neg A]$
- (iii) $v[C \vee (A, B, D)] = v[C] + \frac{1}{2} (v[\neg A] + v[\neg B])$
- (iv) $v[D \vee (A, C, B)] = v[D] + \frac{1}{3} (v[\neg A] + v[\neg C] + v[\neg B])$

Though in and of itself quite neat, Proposal 1½ is poorly motivated. Taking the fact that B 's net value is higher than C 's as a reason for thinking that B is required relative to C simply overlooks A 's availability: given A , B cannot possibly be required. Put differently, at the root of Proposal 1½ lies the view that, relative to C , refraining from B involves a loss of some sort; but given A , refraining from B involves *no* loss.

I accordingly embrace Proposal 2.

11. Association

You will recall that in two-valued logic, as well as in the Łukasiewicz family of many-valued systems, disjunction associates. Thus, $[(p \vee q) \vee r] \leftrightarrow [p \vee (q \vee r)]$. Does dilemmatic disjunction also associate?

To begin with note that (for example) the construction $A \vee (B \vee C)$ reads 'choosing A given choosing B given C ,' which is plainly unintelligible. Therefore, this construction and its like cannot be admitted as wffs. There is then a sense in which the question of association simply doesn't arise.

At the same time, there is also a sense in which the question does arise. $A \vee (B, C)$ for example reads ‘opting for A given B and C .’ And it makes perfect sense to ask whether the value of opting for A given B and C is equivalent to the value of opting for A given C and B . The question before us then is whether $[A \vee (B, C)] \Leftrightarrow [A \vee (C, B)]$. It is of course (I take this to be obvious).

12. The Road Not Taken

In drafts of this Chapter I suggested that a dilemmatic disjunctive act amounts to choosing not *whatever*—but rather the *required*—act among a pair of dilemma-generating actions. Its value, I further suggested, equals the value of that act. In symbols (\vee stands for ‘choosing the required act’).

If $v[A] + v[\neg B] \geq v[B] + v[\neg A]$, then $v[A \vee B] = v[A]$

If $v[B] + v[\neg A] > v[A] + v[\neg B]$, then $v[A \vee B] = v[B]$

This proposal has much to recommend it. In the first place, it entails that dilemmatic disjunction both commutes and associates. Commutes, because if A is required, then both $v[A \vee B]$ and $v[B \vee A]$ equal $v[A]$; which means that $[A \vee B] \Leftrightarrow [B \vee A]$. Associates, because—assuming A ’s net value is higher than B ’s, and B ’s higher than C ’s— $(A \vee B) \vee C$ reads ‘choosing the required among A and B (namely A), and then choosing the required act among A and C (namely, A).’ And $A \vee (B \vee C)$ reads ‘choosing the required act among B and C (namely B), and then choosing the required act among B and A (namely, A).’ It follows that both $v[A \vee (B \vee C)]$ and $v[(A \vee B) \vee C]$ equal $v[A]$. Or that $[(A \vee B) \vee C] \Leftrightarrow [A \vee (B \vee C)]$.

Secondly, according to the suggestion at hand, the value of dilemmatic disjunction is *unconditional*: its value remains constant regardless of the act agents actually pick (because its value equals the value of the act agents *should* pick).

Finally, since choosing the required act is invariably right, the value of dilemmatic disjunction is exclusively positive. This means that all dilemmatic disjunctions are quasi-tautologies.²¹

Why then discard so elegant a thesis? For two reasons. The less serious of them is that the thesis is undefined for failing to choose the required act. The weightier reason is that the thesis conflates the syntax of dilemmatic disjunction (the formation rule) with its semantics (the valuation function). From a metaphysical point of view this means that by contrast to negations of actions, dilemmatic disjunctions cannot be individuated without reference to their deontic status; which is implausible. I mean: perhaps one could motivate (on consequentialist grounds?) the view that *no* moral act can be individuated without reference to its deontic status. But how would one justify this position with respect to only *some* moral acts?

13. Zooming Out

One might worry that these comments open a Pandora's box. The comments led me to maintain that dilemmatic disjunction consists in choosing not the *required*—but rather *whatever*—act among a pair of dilemma-generating actions. At the same time I hold that all dilemma-generating actions are *required*. I am thus committed to the opinion that dilemmatic disjunction consists in choosing whatever act among a pair of

²¹ A quasi-tautology is a formula that never assumes the truth-value 'false' regardless of the truth-values assigned to its propositional variables (Rescher, *Many-valued Logic*, 30).

required actions. It follows that the revised definition too conflates the syntax of dilemmatic disjunction with its semantics.

To meet this objection, let \mathbf{V} read ‘choosing whatever act among a pair of incompatible actions of whatever deontic status.’ And let A and B be a pair of incompatible actions. Then the evaluation procedure for \mathbf{V} in outline is:

- (1) if both A and B are required acts, then...
- (2) if both A and B are forbidden.²² then...
- (3) if A is required while B is forbidden or vice versa, then...

The evaluation function I introduced in Section 7 completes line (1). (This function is thus but a detail of a bigger picture.) What about the evaluation functions completing lines (2) and (3)? (As it turns out, they have a single evaluation function common to both.) Consider first line (2), where A and B are both forbidden acts. There are two cases. First of all, an agent chooses and attempts to perform a forbidden act. Since the agent should have *and could have* refrained from both acts, their incompatibility is simply irrelevant. Choosing to murder is choosing to murder, regardless of whether it’s incompatible with choosing to rob. For all intents and purposes then, $A \mathbf{V} B$ is tantamount to A , and $B \mathbf{V} A$ —to B .

Secondly, the agent refrains—as she can and should—from both A and B . In that case, the agent refrains from A and refrains from B . Thus, $\mathbf{V} A, B$ amounts to nothing but $\neg A \wedge \neg B$.

Similar considerations apply to line (3), where A is required and B forbidden (or vice versa). Since the agent should and can perform A and refrain from B , the fact that

²² Recall that the system under construction is incapable of accommodating super- and suberogatory acts. Thus, an action is required iff it’s right: so if it’s wrong, it’s forbidden.

these acts are incompatible is morally irrelevant. Choosing to save a life is choosing to save a life, regardless of whether it's incompatible with murder. Thus, $A \vee B$ is tantamount to A , and $B \vee A$ —to B . Since refraining from both A and B is possible, $\vee A, B$ amounts to $\neg A \wedge \neg B$.

Accordingly, the evaluation function completing lines (2) and (3) is:

$$(i) \quad v[A \vee B] = v[A]$$

$$(ii) \quad v[B \vee A] = v[B]$$

$$(iii) \quad v[\vee A, B] = v[\neg A \wedge \neg B] = \frac{1}{2} (v[\neg A] + v[\neg B])^{23}$$

Incompatibility of acts then is morally significant when and only when the acts in question are both required. This is the reason why I made line (1) the focus of this Chapter. In other words, I am concerned with offering syntax and semantics for *major* moral connectives. It is only thanks to its application to pairs of required acts that \vee —which reads: choosing among *incompatible* acts—is a major moral connective. Thus, only in its application to pairs of required act does \vee merit careful attention.

The same considerations also explain why I discussed moral dilemmas first and then zoomed out to represent the big picture: they establish in particular that if I discussed 'or' first and then zoomed in on moral dilemmas, the present Chapter would seem poorly motivated.

14. Making It Explicit

Implicit in the foregoing discussion are four major assumptions concerning moral dilemmas. They should be made explicit. (I shall of course make no attempts at defending them here.)

²³ Again, explanation of this line must await until Chapter 4, Sections 9-10.

The first assumption is that refraining from both horns of a moral dilemma is not merely wrong, but also forbidden. Which means that refraining from both horns of a dilemma can never be suberogatory.²⁴ Secondly, to generate a dilemma, required acts need not be equivalent in value.²⁵ Thirdly I assume that all dilemmas are rationally resolvable. For I hold that agents should always aim at maximizing net moral value; and whether an act maximizes net value is (in principle if not in practice) rationally determinable. (Recall that if conflicting acts are equivalent, either one maximizes net value. It therefore matters not which horn agents opt for: either choice would be rational. Thus, dilemmas involving equivalent acts are also rationally resolvable.) Finally, an act (truth-telling for instance) may be required in and of itself, and yet wrong relative to a conflicting act (saving a life); since no wrong acts are required, this means that deontic status is a context-dependent feature of actions.

I should like to say a word on how this account fits into the debate concerning moral dilemmas. The debate centers around a pair of valid arguments, which—taken together—generate a contradiction.

- (1) If agents ought to do *A* and agents ought to do *B*, they ought to do both *A* and *B* (this is the agglomeration principle²⁶).
- (2) I ought to do *A*.
- (3) I ought to do *B*.
- (4) Therefore, I ought to do both *A* and *B*.

- (1) ‘Ought’ implies ‘can’ (or ‘cannot’ implies ‘ought not’).²⁷
- (2) Due to some contingent feature of the world, I cannot do both *A* and *B*.
- (3) Therefore, it’s not the case that I ought to do both *A* and *B*.

²⁴ No author I know of argues to the contrary.

²⁵ Many authors on dilemmas share this assumption. See for example Bernard Williams, “Ethical Consistency,” 125. However, Walter Sinnott-Armstrong believes that dilemma-generating acts must be equivalent in value (*Moral Dilemmas*, 29).

²⁶ See Williams, “Ethical Consistency,” 130.

²⁷ This principle embodies the conviction that a command to do the impossible is not a genuine command.

To avoid this predicament (who is ready to be caught advancing a theory that entails contradictions?), utilitarians²⁸ and deontologists²⁹ alike have argued that their moral systems never impose conflicting duties on agents: if *A* and *B* are conflicting obligations, the theory equips agents with the wherewithal for determining which of them should be acted upon. Take two-valued utilitarianism for example. Suppose that *A* and *B* are conflicting obligations. Now it turns out that *A* maximizes utility. Acting on *B*—failing to maximize utility—would thus be wrong. Since no wrong acts are dutiful, two-valued utilitarianism requires agents to perform *A* alone. (A legitimate worry: if *B* fails to maximize utility, why say it's dutiful in the first place? Though I can think of a way or two around this problem, I shall not discuss them here.)

Bernard Williams on the other hand believes that *B* may on occasion retain its deontic status, even if agents act for the best in choosing *A*.³⁰ (A legitimate worry: if—as Williams supposes³¹—*B* shouldn't be acted upon (because agents should act for the best), how can it remain obligatory?) To handle the problem at hand, he rejects the agglomeration principle.³²

I imagine it's clear that my strategy is more akin to the former than to the latter one.

A final remark. The debate on moral dilemmas is often taken to be revolving around the question whether genuine dilemmas exist. While Williams (and Thomas Nagel,³³ and Walter Sinnott-Armstrong³⁴) answer this question in the affirmative,

²⁸ See the selections from Mill in Gowans, *Moral Dilemmas*.

²⁹ See the selections from Kant in Gowans, *Moral Dilemmas*.

³⁰ "Ethical Consistency."

³¹ *Ibid.*, 122.

³² *Ibid.*, 134.

³³ "The Fragmentation of Value."

³⁴ *Moral Dilemmas*.

utilitarians and deontologists reply to it in the negative. Drawing on this terminology, I in effect assume that there are no genuine dilemmas. But sometime, somewhere, I shall attempt to *demonstrate* this claim. I shall attempt to show in particular that the following *modus tollens* is sound.

- (1) If there are genuine moral dilemmas, then rightness and wrongness aren't scalar properties.
- (2) But rightness and wrongness *are* scalar properties.
- (3) Therefore, there are no genuine dilemmas.

15. Summary

A disjunctive action amounts to choosing among a pair of incompatible acts, and attempting to act upon the choice. Where A and B are both required, the evaluation procedure for \mathbf{V} is:

If $v[A] + v[\neg B] > v[B] + v[\neg A]$, then

- (iii) $v[A \mathbf{V} B] = v[A]$
- (iv) $v[B \mathbf{V} A] = v[B] + v[\neg A]$

If $v[A] + v[\neg B] < v[B] + v[\neg A]$, then

- (iii) $v[A \mathbf{V} B] = v[A] + v[\neg B]$
- (iv) $v[B \mathbf{V} A] = v[B]$

If $v[A] + v[\neg B] = v[B] + v[\neg A]$, then

- (iii) $v[A \mathbf{V} B] = v[A]$
- (iv) $v[B \mathbf{V} A] = v[A]$

If $v[\neg A] > v[\neg B]$, then

- (i) $v[\mathbf{V} A, B] = v[\neg A]$

If $v[\neg B] > v[\neg A]$, then

- (i) $v[\mathbf{V} A, B] = v[\neg B]$

Chapter 4: 'And'

1. Fixing Ideas (1)

In two-valued logic 'and' is definable in terms of inclusive 'or' and 'not': $A \wedge B$ is equivalent to $\neg(\neg A \vee \neg B)$.¹ In Łukasiewicz's many-valued logic too, $A \wedge B$ is equivalent to $\neg(\neg A \vee \neg B)$.² (In Łukasiewicz's system, disjunction, conjunction, and negation are (respectively) defined as follows (where $v[p]$ is p 's degree of truth, which we suppose is a real number between 0 and 1):

$v[p \vee q] = \max \{v[p], v[q]\}$ (the degree of truth of a disjunction is the maximum of the degrees of truth of its disjuncts)

$v[p \wedge q] = \min \{v[p], v[q]\}$ (the degree of truth of a conjunction is the minimum of the degrees of truth of its conjuncts)

$v[\neg p] = 1 - v[p]$

Thus, if for example $v[p] = .3$ and $v[q] = .8$, then $v[p \wedge q] = .3$, $v[\neg p] = .7$ and $v[\neg q] = .2$, which means that $v[\neg p \vee \neg q] = .7$, which in turn means that $v[\neg(\neg p \vee \neg q)] = .3$. Thus, $v[p \wedge q] = v[\neg(\neg p \vee \neg q)] = .3$.)

Is 'and' in the present system likewise definable in terms of 'or' and 'not'? To answer this question, it is necessary to first articulate what a conjunction of actions amounts to. I shall suggest that it amounts to performing more than one act either simultaneously or at different times.

¹ Similarly, disjunction is definable by means of conjunction and negation: $A \vee B$ is equivalent to $\neg(\neg A \wedge \neg B)$.

² And $A \vee B$ is equivalent to $\neg(\neg A \wedge \neg B)$.

2. Fixing Ideas (2)

To say that Sharon is conjoining changing the baby and granting phone sex services³ presumably means that she is performing both actions (or courses of action) simultaneously. ‘Both’ because the agent carries out *two*—that is, numerically distinct—actions: meaning that neither of her acts somehow amounts to the other: changing the baby doesn’t in any shape or form amount to granting phone sex services, nor does granting phone sex services amount to changing the baby. By contrast, Brutus’s stabbing Caesar does amount to his killing of Caesar, because the killing was committed *by* stabbing. Sharon’s acts then are distinct from each other in the sense that neither of them can be seen as a part or re-description of the other. Actions that *can* be so seen are the (difficult) topic of Chapter 5.⁴

Generalizing from this example I propose that a conjunction of n ($n > 1$) actions consists in performing n distinct actions at once. (In theory, the only limitation on n ’s size is that it be bigger than 1; in practice, its size rarely exceeds 3. This is because our attention resources are fairly limited, as is the number of bodily parts (such as the limbs) with which we perform what can be legitimately called ‘actions.’) Thus, while negation of an action amounts to refraining from it, and while disjunction of actions amounts to choosing one of two incompatible acts (both of which are called for), a conjunction of actions (or a conjunctive act) amounts to carrying out more than a single action simultaneously.

³ This example is drawn from a scene in Robert Altman’s *Shortcuts*, a film based on Raymond Carver’s short stories.

⁴ Alvin Goldman proposes a criterion of identification for simultaneous and independent actions (he calls the latter ‘co-temporal’): the correctness of saying that an agent is doing one of the acts *while also* doing the other (*A Theory of Human Action*, 22.) Thus, ‘Sharon is changing the baby while also granting phone sex services’ is correct, but ‘Brutus was stabbing Caesar while also killing him’ isn’t.

To be sure, for actions to be simultaneous, they need not *begin and end at the same time*: after all, smoking and driving may be performed simultaneously even if they overlap only partially (say, the former begins after the latter was initiated and comes to an end before the latter has). Rather, whenever and *for as long as* an agent carries out at least two distinct actions, she performs a conjunctive act.

3. Simultaneity, Characters and Morality

One might complain that my proposal has an air of paradox. For on the one hand it states that conjunction of actions consists in performing at least *two distinct acts* (simultaneously); on the other—that it is a *single act* (namely, the conjunctive act). On a closer look however my view turns out to be no more paradoxical than any accepted interpretation of conjunction. A conjunction of propositions (or states of affairs) is after all a single proposition consisting in two distinct propositions (each of which may of course consist of further propositions). Ditto with respect to conjunction of sentences. Since the paradox involved in conjunction of propositions or sentences generates no particular discomfort, the paradox involved in conjunction of *acts* shouldn't either.

What *should* cause unease about the present proposal is its insistence on simultaneity. Isn't cooking and cleaning up *afterwards* a conjunction of two courses of action? Or exercising and *then* taking a shower? Why deny that conjunction consists in carrying out several distinct actions *at different times*?

Not only there is no reason for denying this thesis, there is an important reason for embracing it. Moral acts that a person executes at different times, particularly over extended time spans (significant chunks of a life), are of special moral interest. These acts form a *pattern* of behavior, thus providing spectators with a solid ground for judging

that a person is generous, industrious, mean, condescending, and in general good or bad. In short, these acts serve as a basis for *character* evaluations (concurrent moral acts are simply too rare to serve for character assessments).

A person's character in turn is morally significant in various ways. In the first place (and as already indicated), characters are a subject of moral evaluation: they are said to be virtuous or vicious, just as social institutions and policies are said to be just or unjust and actions right or wrong. Moreover, character often affects the moral value of acts, a phenomenon most conspicuously visible in criminal judicial proceedings. Recidivism for example normally serves as an aggravating factor in the courtroom, especially if the previous offenses are similar to the one for which the offender is being sentenced. One justification for this policy (which stems from the view that the purpose of punishment is retribution) is that if sentences symbolize moral disapproval not only of the offense but also of the offender (i.e., of his character), a person who has committed only one offence can—often if not always—claim that it was 'out of character,' and thus benefit from mitigation; but this claim becomes less plausible the richer one's record of previous convictions gets.⁵ On the other hand, evidence of good character—such as bravery—is sometimes allowed as a mitigating factor *even when it has nothing to do with the offence*. (This policy too seems retributive in spirit.) Thus, people have had prison terms shortened because they have fought well in a war, given a kidney to a sister, saved a child from drowning or started a youth club. In such cases meritorious conduct is taken

⁵ See Walker, *Sentencing: Theory, Law and Practice*, 44.

as evidence that the offender is a better and more valuable member of society than her offence suggests.⁶

Retributivism is but one explanation of why character affects moral value. Another account (call it ‘agent-based’; clarification is forthcoming) appeals to motives. Since the motives of a virtuous individual are good, his actions are right even when their consequences are bad. Likewise (*mutatis mutandis*) with regards to vicious individuals. On this account then character affects moral value via influencing assessments of motives.

Note. While the foregoing discussion implies that character serves to determine the moral value of acts *and vice versa*, I don’t wish to suggest that character and moral value *interact*. The moral value of acts determines character in the epistemic sense: I take it that a person’s actions are external and observable manifestations of his inner and unobserved—perhaps unobservable—personality traits; thus, actions serve to determine one’s character in the sense that they reveal it.⁷ On the other hand, character determines moral value in neither an epistemic nor a causal sense, but rather a definitional or conceptual sense: good character (according to the agent-based account) is but a criterion of rightness (an act is right iff the agent is virtuous).

⁶ Walker believes that this policy results from two assumptions: (i) that offenders are being sentenced not for the offence but for their moral worth, and (ii) that moral worth can be calculated by a sort of moral book-keeping (*ibid.*, 50).

⁷ Recall however that according to Aristotle, actions determine character in the *causal* sense: by doing the acts that we do in our transactions with other men we become just or unjust, and by doing the acts that we do in the presence of danger, and by being habituated to feel fear or confidence, we become brave or cowardly. . . . Thus, in one word, states of character arise out of like activities.” (*Nicomachean Ethics*, Book II, Chapter I).

I say more on this notion later on.

4. Digression: Note on Virtue Ethics

Talk of character and morality brings to mind virtue ethics. The foregoing remarks are meant to neither faithfully represent nor constitute a version of virtue ethics (for one thing, I have no grip of what virtue ethics *is*: I take special (though admittedly somewhat infantile) pleasure in belaboring this confession). Yet examining them in light of Michael Slote's taxonomy of the various approaches to virtue ethics⁸ suggests that they bear affinities to Plato's view (Slote dubs it 'agent-based') as well as to Aristotle's ('agent-focused').

According to Slote, Aristotle maintains that the virtuous individual accurately perceives what ought to be done in a given situation, and acts upon his perception rather effortlessly.⁹ Plato (according to Slote again) holds that the motives or inner states of the virtuous agent serve to *make* his actions morally right.¹⁰ For Aristotle then the moral value of acts is independent of the agent's inner states: for Plato, the morality of actions is derivative from the individual's inner states or motives. Since I too believe that an action's rightness derives from the agent's motives, my position is similar to Plato's. At the same time I hold that actions reveal a person's character: which suggests that moral value is independent of a person's inner states.

How could one consistently endorse both Plato's and Aristotle's standpoints? By assuming that well-intended acts normally have good consequences, and ill-intended acts—bad consequences. (What about the *klutz* whose actions have invariably bad

⁸ "Virtue Ethics."

⁹ *Ibid.*, 178.

¹⁰ *Ibid.*, 178, 206-16.

consequences even though all are well-intended? The Talmud and I urge him to avoid acting altogether.) A different type of response is to be found in yet a third approach to virtue ethics, which Slote names ‘agent-prior’ and attributes to Rosalind Hursthouse. It states that although act evaluations are derivative from evaluations of character traits and motives (as the agent-based position has it), evaluations of traits and motives are not fundamental but instead are based in further ethical facts about human well-being and flourishing (as the agent-focused approach has it). In other words, the agent-prior position states that acts are right if they accord with or exercise virtue, and a trait counts as a virtue if people need it in order to flourish or have good lives.¹¹

5. Still Fixing Ideas

I argued that an agent’s personality is morally significant. And also that it is predominantly the collection of behaviors that a person exhibits *at different times* (particularly over extended segments of time) which serves to determine her character. It stands to reason therefore that conjunction be applied to character-revealing behaviors. Yet I apply it to behaviors (concurrent moral acts) that—as far as personalities are concerned—are hardly informative.

Explaining this choice is but one challenge to my proposal (namely, that conjunction of actions consists in performing more than a single act simultaneously). Another is to show how a person’s character can nonetheless be taken into account in the present system (as it should, given its significance).

It appears however that there is little point in embarking on an attempt to meet these challenges. Instead it should be acknowledged that there is ample reason for replacing my initial proposal (‘synchronic’) by another: that conjunction of actions

¹¹ *Ibid.*, 207.

consists in performing n distinct actions *at different times* ('diachronic'). Also possible is a unifying thesis which states that conjunction of actions amounts to performing n distinct actions *either simultaneously or at different times*. Indeed, such a reading makes much sense, as concurrent moral acts are not *uninformative* with respect to character: there is therefore no reason to exclude them should they occur. I shall accordingly do just that: replace the synchronic by the unifying interpretation.

6. Actions are Things

One might wonder how a conjunction of actions could itself be an act, given that there is no additional action (bodily movement(s)) over and above the actions that an agent carries out either concurrently or at different times. This worry will be relieved once it is acknowledged that a conjunctive act is nothing but the mereological sum of the actions which compose it as parts (i.e., its conjuncts).

The mereological sum (or fusion) of several *things* (which is the least inclusive thing that includes all of them as parts) is composed of them and of nothing more.¹² Now there are compelling reasons for taking actions to be things (or values of variables).¹³ One such reason is that it explains certain entailment relations, such as those involving adverbial modification.¹⁴ In particular, if

(1) Jones buttered the toast

¹² This characterization is based on David Lewis's (*The Plurality of Worlds*, 69 fn.), which is in turn based on Henry Leonard and Nelson Goodman's work on the topic ("The Calculus of Individuals and Its Uses."

¹³ The view that events and actions (the latter are often seen as a species of the former, and I shall follow suit: note however, that some authors—for instance, Kent Bach ("Actions are not Events")—have raised considerations against this position)—the view then that actions and events are things has a long and respectable history. (For details, see Roberto Casati and Achille Varzi's introduction to their *Events*.) But it is probably Davidson who—through his "The Logical Form of Action Sentences"—introduced this view to current philosophical discussion.

¹⁴ The reason is due to Davidson (*ibid.*). However, it was Anthony Kenny (*Action, Emotion and Will*) who pointed out the problem of accounting for these logical relations.

is analyzed as consisting of a binary predicate (i.e., ‘battered’) flanked by two singular terms (i.e., ‘Jones’ and ‘the toast’), it’s unclear why it is entailed by

(2) Jones battered the toast with a knife

in which ‘battered with’ is construed as a distinct three-place predicate. If however (1) is analyzed as consisting of a three-place predicate with a bound action variable—that is, if (1) is analyzed as follows

$(\exists a) (\text{Buttering} (\text{Jones, the toast, } a))$

and (2) is analyzed as follows

$(\exists a) \{ \text{Buttering} (\text{Jones, the toast, } a) \wedge (\exists b) [\text{knife} (b) \wedge \text{with} (a, b)] \}$

(so that the adverbial expression ‘with a knife’ is thought of as a simple predication of an action), then it’s clear why (2) entails (1).

Like any substantial theory, the theory that actions are things hasn’t escaped the pointed criticism of some shrewd thinker or other. Terence Horgan for example argues that there are means for handling issues such as the logic of adverbial modification, which do not require quantification over actions and events (that is, means that don’t require taking actions and events to be *bona fide* entities). Specifically with respect to adverbial modification Horgan suggests that adverbs are operators added to an ordinary first-order predicate calculus. Syntactically these operators precede well-formed formulae, forming more complex well-formed formulae; semantically they can be construed as functions which map the properties expressed by the formulae they modify onto new properties.¹⁵ Horgan concludes that events and actions may be eliminated from

¹⁵ This suggestion has been advanced by several authors independently of each other. See for example Terence Parsons, “Some Problems Concerning the Logic of Grammatical Modifiers.”

our ontology, a move recommended by Occam's Razor.¹⁶

Yet it's far from clear that eliminativists respect theoretical parsimony more than friends of actions and events. As Altman, Bradie, and Miller argue,¹⁷ while the latter must proliferate entities in the domain, the former are forced to proliferate logical connectives (they must admit special, nontruth-functional connectives). Therefore, eliminativism isn't obviously preferable over the view that actions and events are things: the latter's plausibility remains unshaken.

I accordingly take it that actions are things. This being so, the mereological sum of several *actions*—like that of several things—is made up of these actions and of nothing more. The mereological sum of several things constitutes an additional thing without thereby requiring extra matter: an assembled car doesn't contain more matter than the sum of its parts. (In connection with this example Professor Rohit Parikh has remarked that if actions are things, an assembled car consists of the sum of its part plus the action of assembling it; which means that an assembled car contains more things than an unassembled one. Well, in the first place, actions and events are things only for as long as they occur; for once they cease, they no longer occupy some portion of space-time. Since an already assembled car is no longer *being assembled*, it doesn't contain more things than an unassembled car. Secondly, holding that actions and events are part of the furniture of the world doesn't entail that they and material objects fall into the same metaphysical category. While I cannot here delve into the question of what distinguishes events from material objects (it would require a book discussion), insisting that they inhabit different metaphysical categories would allow me to maintain my

¹⁶ "The Case Against Events."

¹⁷ "On Doing Without Events."

original claim: namely, that an assembled car doesn't contain more matter (or more material objects) than the sum of its parts.)

I said then that the mereological sum of several *things* constitutes an additional thing without thereby requiring extra matter. Likewise, the mereological sum of several *actions* adds up to an additional action—the conjunctive act—without thereby requiring extra bodily movements. Thus, whenever an agent performs n distinct actions concurrently, she *does* in fact perform $n + 1$ actions: but the $n + 1^{\text{st}}$ action (the conjunctive act) includes no bodily movements over and above those required for executing the conjuncts.

As to (i) why several *distinct* moral actions add up to an additional *entity* (which suggests *uniformity*), and (ii) why the mereological sum of distinct moral actions is itself an action, I take up these problems next.

7. Problem (i)

Since (I take it) actions are things or individuals, I apply to them the calculus of individuals developed by Leonard and Goodman.¹⁸ A postulate of the calculus states that every two individuals have a sum which is itself an individual.¹⁹ (This individual and any other individual also have a sum which is itself an individual; ultimately then any number of individuals have a sum which is itself an individual). It follows that any two acts have a sum which is itself an individual (but ultimately, the sum or conjunction of any number of actions is itself an individual).

¹⁸ See "The Calculus of Individuals and Its Uses." The calculus is also outlined in Goodman's *The Structure of Appearance*. References are to the latter.

¹⁹ *Ibid.*, 36. As already noted, Leonard and Goodman define the sum of two individuals as the least inclusive individual that includes both of them as parts (*ibid.*, *ibid.*). (The mereological sum of any number of individuals is the least inclusive individual that includes all of them as parts.)

As Goodman notes, the statement that *every* two individuals have a sum which is itself an individual tends to raise eyebrows. For (it is argued) the mereological sum of several things sometimes constitutes an additional thing, but doesn't on other occasions. Consider first one and the same collection of things. The parts of which a car is composed add up to an extra object—a car—given a certain arrangement; in its absence, the parts fail to constitute a car. (Of course, the parts may fail to constitute a car even in *partial* absence of the required arrangement; for present purposes however there is no need to insist on or pursue this claim). Now consider different collections of things. A car's components (appropriately arranged) add up to an extra thing, whereas Caesar's nose and the state of Utah presumably don't.²⁰ But if *every* two individuals have a sum which is itself an individual, then any two very different and remote individuals—such as the Arctic sea and a speck of dust in the Sahara or Caesar's nose and the state of Utah—have a sum that is an individual. Which is absurd.²¹ The worry then is but an expression of the sentiment that individuality implies some minimum of coherence, organization, homogeneity, uniformity, or continuity of boundaries.

This objection pertains to my position as well: for why think that two very remote actions of the same agent—say, actions of his when he is three and forty years old—constitute an additional thing? Why do *distinct* actions add up to an *entity* or individual?

Goodman responds that the objection reflects a misunderstanding of the range of the calculus' individuals. Contrary to common intuitions, an individual need *not* be organized, not even minimally organized. Likewise it need *not* be uniform, have

²⁰ Clearly, one and the same collection of things may make up two different extra objects (at different times). For instance, one and the same collection of building blocks may constitute a tower and a bridge. This case is however of no relevance for present purposes.

²¹ *Ibid.*, 36-7.

continuous boundaries, or the like. The technical use of the term 'individual' must be freed of certain restrictive associations arising from popular usage. just as the technical use of 'class' is freed from such associations (no one who uses 'class' for technical purposes supposes that several people must take instruction together, or be of similar social position, or have some other special characteristic in common, in order to constitute a class.)

Goodman's response serves my goals well. Yet it's worth noting that my position doesn't *entirely* betray the intuition that individuality implies uniformity. This is because the distinct actions performed by an agent over an extended time stretch *do* have a common denominator: very roughly, they all originate from the same person (this is why they inform spectators about an individual's *personality*). 'Very roughly,' because an agent's actions when she is three and thirty years old *don't* originate from the *same* person: personalities change over *such* time spans.²² The time segment in question then should be long enough to allow evidence about an individual's character to accumulate: at the same time, it shouldn't be long enough to allow the agent's personality to have changed dramatically.

8. Problem (ii)

I have established—and thereby handled problem (i)—that a conjunctive act is an individual. This individual may or may not be an action (the sum of two persons is after

²² It follows that 'same' isn't transitive. In particular, a person *P* at a time t_0 is presumably the same as *P* one moment later, two moments later, and so on. Proceeding in this way long enough entails that *P* at t_0 is the same as *P* twenty years later. Since I deny this conclusion, I must concede that 'same' isn't transitive (for even though *P* at t_0 is the same as *P* one moment later, two moments later.... it is not the same as *P* twenty after t_0). Isn't the *sorites* paradoxical. (Note however that in at least some contexts the paradox may be softened by replacing 'the same' with 'similar'.)

all an individual but not a person). I now turn to argue that it is, thus taking care of problem (ii).

In performing several distinct generous acts, an agent reveals or exercises her generosity. In performing several callous acts, an agent manifests or exercises her callousness. (It is thanks to their consistent recurrence or stability that acts may be considered a manifestation of a *trait*. More on this matter shortly.) And in performing several right (or wrong) acts of various kinds, a person reveals or exercises her virtuous (or vicious) character. (Again, it is due to their persistence that her acts amount to exercising that long-lasting feature called ‘character’).²³ In short, the mereological sum of distinct moral actions amounts to *revealing or exercising one’s moral character (one’s moral traits: virtues and vices)*.

Exercising one’s moral character involves no bodily movements over and above the ones required for carrying out individual moral acts. Yet it is not identical with any individual action, because it involves multiplicity (at least two acts must be performed for an act of exercising one’s moral character to occur). Also, like many actions, revealing one’s moral character is often *intentional*: on many occasions (this is especially so when one assumes an authoritative role) agents intend that the *collection* of their actions—their conduct in its entirety—would reflect a certain moral self; that is, agents form intentions with respect to a *collection* or sum of moral actions. Thus, they consider what type of conduct is likely to portray the desirable moral image and weigh how individual actions

²³ A consequence of this position is that in performing both right and wrong acts in an a-systematic order, a person reveals that she is neither reliably good nor reliably bad or lacks moral character. Insofar as a person also lacks moral *principles* or *theory*—and it would seem that *only* a morally inconsistent agent may also lack principles, for presumably following rules guarantees a considerable degree of consistency—he also lacks moral backbone. (This means that a consistently evil person may possess moral backbone.)

cohere with their intentions concerning the whole (collection).

Now I said that a conjunctive act—exercising one’s moral character—amounts to performing at least *two* distinct actions. But this is implausible, for in performing merely two acts an agent hasn’t yet established that it is *character*—a stable component of one’s self—which is responsible for her conduct (as opposed to sheer good or bad luck). As already noted,²⁴ and as will be further noted later on,²⁵ I cannot agree more. However, it remains true that in performing two (say) generous actions, an agent already *suggests* that she is generous. For two acts, but not a single one, already indicate a *trend*.

Now it wouldn’t be absurd to follow in Aristotle’s footsteps and maintain that a conjunctive action amounts not only to exercising one’s moral character, but also to *shaping or constituting one’s moral self* (an action by means of which one undergoes the process of becoming a person of a certain moral character).²⁶ For Aristotle may be right in suggesting that repeating moral behaviors makes them habitual, thus turning them into these stable and long-lasting features we call ‘personality’ and ‘personality trait.’^{27 28}

Like exercising one’s moral character, molding it involves no bodily movements beyond the ones required for performing individual acts (none of which is identical to the molding itself). Also, like many actions, it involves a *change* in an object (the person), since through it a person transforms from (say) being neutral to being good or bad, from

²⁴ At the end of Section 6.

²⁵ In Section 12.

²⁶ See fn. 7.

²⁷ *Nicomachean Ethics*, Book II, Chapter 1.

²⁸ A qualm. If a person acts generously, she must be generous to begin with. But then, her character determines her acts and not vice versa, contrary to Aristotle’s claim.

Well, Aristotle distinguishes between a case where we would call someone generous because he did a generous deed and the case where we would call someone generous because he has a generous character. This distinction paves the way for holding that someone may choose to perform an act characteristic of a generous man even if he does not yet have the habit of generosity (*ibid.*, Chapter 4).

being good to being better, from being bad to being good, and so on. Moreover, molding one's moral self seems *intentional*, for agents must—at some point or other—form intentions with respect to the *collection* of actions through which they would eventually turn into a person of a certain kind.

Arguably however, molding one's moral character by actions excludes revealing it by them, as the former implies that there is no character prior to actions, while the latter—that there *is* (actions merely reveal it). For this reason I abandon the proposal that a conjunctive action amounts to molding one's moral self.

9. Truth-Table

Having taken care of some worries and sources of discomfort, the ground is now prepared for looking into the issue of what the value of a conjunctive act would be. Given the foregoing discussion, assigning value to a conjunctive act amounts in effect to assigning value to a person's moral character, or to offering a measure of an individual's global moral ability. Observing how another human ability—scholastic aptitude—has been traditionally measured might therefore be of help. One's overall academic ability is indicated by the grade average one has obtained over a sufficiently substantial academic history (the grade average of two or three courses is clearly an inadequate basis for evaluating one's *overall* scholastic aptitude). I propose that a person's moral character (one's *overall* moral ability) is likewise indicated by the average value of a sufficiently large number of actions.

Thus, I propose that the value of a conjunctive act is the *average of the values of its conjuncts*. Following is (a segment of) the truth-table for conjunction:

A_1	A_2	A_3	A_4	A_5	A_6	$A_1 \wedge A_2 \wedge A_3 \wedge A_4 \wedge A_5 \wedge A_6$
.1	.2	.3	.4	.5	.6	$1/6 (.1 + .2 + .3 + .4 + .5 + .6) = .35$
-.8	.1	-.7	-.9	.1	.1	$1/6 (-.8 + .1 - .7 - .9 + .1 + .1) = -.35$
.8	-.1	-.1	-.1	-.1	-.1	$1/6 (.8 - .1 - .1 - .1 - .1 - .1) = .05$

Cast more generally, the truth-table for conjunction is:

$$v[A_1 \wedge A_2 \wedge \dots \wedge A_n] = 1/n (v[A_1] + v[A_2] + \dots + v[A_n])$$

Cast more succinctly, it is:

$$\sum_{i=1}^n A_i / n$$

10. Association

It may seem that, contrary to standard notions of conjunction, the present one fails to associate. This is because (to take an example) $v[(A_1 \wedge A_2) \wedge A_3] = \frac{1}{2} \{[\frac{1}{2} (v[A_1] + v[A_2])] + v[A_3]\}$, whereas $v[A_1 \wedge (A_2 \wedge A_3)] = \frac{1}{2} \{v[A_1] + [\frac{1}{2} (v[A_2] + v[A_3])]\}$ which would normally be a different value. For example, $\frac{1}{2} [\frac{1}{2} (.3 + .4) + .5] = .425$, whereas $\frac{1}{2} [.3 + \frac{1}{2} (.4 + .5)] = .375$. To eliminate these differences (that the differences *ought* to be eliminated I take to be obvious), the value of conjunction must be calculated using the following formula:

$$v[A_1 \wedge A_2 \wedge \dots \wedge A_n] = 1/n (v[A_1] + v[A_2] + \dots + v[A_n])$$

Consequently it might be thought that—contrary to standard conjunction, which always consists of exactly two conjuncts (each of which may of course consist of further conjuncts)—the present conjunction contains as many conjuncts as the actions comprising it.

This however need not be the case. Conjunction would always consist of exactly two conjuncts, provided that its value be the *weighted*—not the *simple*—average of the

values of its conjuncts. Thus, $v[(A_1 \wedge A_2) \wedge A_3] = 1/3\{2 \times [1/2 (v[A_1] + v[A_2])] + v[A_3]\}$

which (to take the same action values as above) equals $1/3\{2 \times [1/2 (.3 + .4)] + .5\} = .4$.

And $v[A_1 \wedge (A_2 \wedge A_3)] = 1/3\{v[A_1] + 2 \times [1/2 (v[A_2] + v[A_3])]\}$ which again equals $1/3\{.3 + 2 \times [1/2 (.4 + .5)]\} = .4$ (which in turn equals $1/3 (.3 + .4 + .5) = .4$ or $1/n (v[A_1] + v[A_2] + \dots + v[A_n])$).²⁹ In short, conjunction always consists in exactly two conjuncts if its truth-table is

$$v[(A_1 \wedge A_2) \wedge \dots \wedge A_n] = n_i \bar{x}_i / (n_i + n_j) + n_j \bar{x}_j / (n_i + n_j)$$

where n_i and n_j are the numbers of values (i.e., the numbers of actions) in the ultimate conjuncts of a given conjunctive act, and \bar{x}_i and \bar{x}_j are the averages of these values.

Moreover, if this is its truth-table, conjunction also associates, as is evident by the fact that $v[(A_1 \wedge A_2) \wedge A_3]$ now equals $v[A_1 \wedge (A_2 \wedge A_3)]$.

11. Commutation

Conjunction then not only always consists of exactly two conjuncts, it also associates, just as standard conjunction does. Just as standard conjunction again, conjunction commutes: $A_1 \wedge A_2$ is equivalent to $A_2 \wedge A_1$, $(A_1 \wedge A_2) \wedge A_3$ to $A_3 \wedge (A_1 \wedge A_2)$, and so on.

12. Character as a Threshold Concept

I said that a person's moral character is indicated by the average value of a sufficiently large number of actions. This means that character trait (like academic ability) is a threshold concept. We are—or at least *should* be—reluctant to infer that a

²⁹ Clearly, the formulae for calculating weighted and simple average produce (as they should) the same value. But only the weighted average allows conjunction to always consist in exactly two conjuncts: which is my main (though not sole: see below as well as Section 13) reason for employing it.

person is (say) generous or cowardly unless he has performed so many acts of generosity or cowardice; that is, unless a certain threshold has been crossed.

More generally, we hesitate to judge that a person is virtuous or vicious prior to having observed a sufficiently large number of right or wrong actions on his part. With one exception though: unusually right or wrong actions weigh more (just as an unusually demanding academic assignment is oftentimes assigned a greater weight): thus, a smaller number—even two—of them should suffice for making character judgments.

The concept of character trait then resembles the concept of speeding in the sense that in order for it to apply, a certain threshold must be crossed. One doesn't speed unless she crosses a certain threshold. Likewise, a person isn't virtuous or vicious unless he meets a certain criterion. As with speeding where the threshold may vary—e.g., it's normally lower for trucks—so with character traits: the threshold is lower for unusually right or wrong actions.

13. Weighted Average, Standard Deviation and Trend

Since the value of a conjunctive act is the weighted average of the values of its conjuncts, facts about weighted average in general—as well as facts about related measures—apply to conjunction. These facts correspond to elements of moral evaluation. Following are four important examples.

Outliers The effect of outliers evaporates. Consider for example a conjunctive act consisting of a single action of value $-.8$ and of five acts of value $.1$. Its value is $1/6 (-.3) = -.05$. Thus, the effect of an action of value $-.8$ is greatly diminished in a neighborhood of five actions of value $.1$. And this is just as well: it coheres well with our

intuitions regarding both academic and moral abilities that outliers shouldn't greatly influence the overall picture (and should perhaps be attributed to good or bad luck).

Partial Information Suppose that you know the value of a given conjunctive action A as well as the number of actions that make up each of its conjuncts, but not the values of these actions. You would now like to take into account an additional action B (an action dug out from the agent's remote past, or a newly performed act), in order to paint a richer picture of the agent's moral character. That is, you are interested in finding out the value of a new conjunctive act, that resulting from A and B . The truth-table for conjunction clearly allows you to achieve this goal, since it doesn't rely on the values of the individual actions that comprise A .

Standard Deviation The standard deviation of a distribution indicates how homogenous it is: that is, how consistent a given phenomenon is.³⁰ When the phenomenon in question is moral behavior, the standard deviation discloses how consistent a person is with respect to a certain trait: how consistently kind or rude she is, for instance. Or with respect to moral character in general: how consistently virtuous or vicious he is.

The standard deviation also allows to observe and express moral differences between people that would otherwise seem identical. Consider the following (samples of) distributions.

Distribution I: .3, .3, .3, .3, .3, .6

³⁰ Where x_i is a value in the distribution, \bar{x} is the distribution's average, and n the number of values in the distribution, the standard deviation is:

$$SD = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Distribution II: .1, .2, .3, .4, .5, .6

These distributions yield the same average value (.35) and yet they differ considerably in their homogeneity: the standard deviation of Distribution I is .122, whereas that of Distribution II is .187.³¹ Average values then fail to reflect the moral difference between their respective sources (which—I am assuming—are two different individuals). The standard deviation by contrast conveys the difference: it indicates that the person responsible for Distribution I is (in all likelihood) more consistently good than the person whose conduct is registered by Distribution II.

Trend Over a period of two years, Mendel and Yankel produce the following distributions, respectively: 1, .2, .3, and .4, and .4, .3, .2, and .1. These equally large distributions are identical in both average and standard deviation, yet Mendel and Yankel differ in a morally significant way. While Mendel is improving (and hence should be encouraged), Yankel is on the decline (and should be discouraged). The trend of an individual's moral ability can be detected by analysis of variance with contrasts for linear trends (I leave out the technical details), or simply by "eyeballing."

14. Objection: Misleading Equalities and the Interval

As far as their resources (time, finances, upbringing and the like) are concerned, Moishe and Avrum are equally equipped to commit morally right actions. Yet while Moishe performs ten actions of value .7 each, Avrum performs twenty such acts. The truth-table for conjunction implies that Moishe and Avrum have an equally good character ($1/10 (.7 \times 10) = 1/20 (.7 \times 20) = .7$). But surely, Avrum is a better person than

³¹ This example is clearly concerned with comparing two equivalent conjunctive actions, each of which consists of the same number of conjuncts. Comparisons of equivalent and unequal conjunctions will be discussed below.

Moishe.

Well, comparing agents requires equality in *all* relevant respects: that is, it requires a *ceteris paribus* clause. Thus, concluding that Avrum is a better person than Moishe requires assuming not only that they are equally equipped to perform right acts, but also that they have committed their respective acts over equivalent time stretches. (To see this more clearly suppose that Moishe has obtained a certain grade average for eight courses over a period of one year, whereas Avrum has obtained roughly the same grade average for sixteen courses over a period of two years. Under these circumstances we may legitimately infer that Moishe and Avrum possess the same academic ability. But if Moishe and Avrum have built up the said academic records over the same period of time, it would be reasonable to conclude that Avrum is academically stronger.)

Alas, the truth-table for conjunction entails that Moishe and Avrum are equally virtuous even assuming that they carry out their respective acts over equivalent time spans. The truth-table in question then fails to represent a material piece of moral information.

It appears that this kind of information could only be captured by comparing the totality of moral value generated by each agent: $10 \times .7$ or 7 in Moishe's case; $20 \times .7$ or 14 in Avrum's. Trouble is, 7 and 14 lie outside the system's interval (the interval between -1 and 1). For totalities of moral value to fall inside the interval, the latter must be open-ended.

However, there is a powerful reason for preferring the system's interval over an open-ended one: the latter would undermine negation and consequently disjunction.

Recall that the truth-table for negation is:

1. If $-.5 \leq v[A] \leq .5$, $v[\neg A] = -1.7 v[A]$
2. If $v[A] > .5$, $v[\neg A] = -1.7(1-v[A])$
3. If $v[A] < -.5$, $v[\neg A] = -1.7(1+v[A])$

It reflects the view that negation values of acts in lines 2 and 3 should mirror negation values of acts in line 1. (You will recall that the intuitions propelling this view concern the inter-relations between, on the one hand, the cause for which an act is committed and, on the other hand, the sacrifice an act involves for the agent.) But capturing this view is impossible if the interval is open-ended. The reason is that such an interval has no middle point that can serve—as .5 does—as a “turning point.” (Naming an arbitrary “turning point”—say 4—might at first blush seem an attractive answer to this problem. As a matter of fact however, it entails that infinitely many values—any value higher than 8—have no negation value whatsoever.)

If an open-ended interval is out of the question, how would the problem at hand (i.e., that conjunction entails that Moishe and Avrum are equally virtuous even though they aren't) be handled? By recalling that avoiding an opportunity to produce rightness (refraining from a right action) is wrong. Specifically, since they are equal in *all* relevant respects, Moishe and Avrum have equal opportunity to produce rightness. If Moishe nonetheless produces less total rightness than Avrum, it must be because he avoids some opportunities to generate rightness (he refrains from performing right acts). Since such conduct is wrong,³² his measure of moral ability (the average value of his actions) would *differ* from Avrum's. On the other hand, if Avrum produces more total rightness than Moishe because his life accidentally presents him with more opportunities to carry out right acts, concluding that the agents are equally virtuous makes perfect sense. After all,

³² Recall that supererogatory acts are beyond the system's scope.

it stands to reason that Moishe *would* produce as much total rightness as Avrum had he only had a chance.

In sum, if Avrum is better than Moishe, conjunction doesn't entail that they are equally good: and if they are equally good, conjunction entails that they are. Either way then, conjunction delivers an intuitive judgment.

An alternative treatment to the problem at hand consists in deploying the statistical tool known as Student's *t*-test. In one of its two applications, the *t*-test is a method of testing whether the mean of a small sample drawn from a normally distributed population is significantly different from the true (population) mean. The test is so structured that as the sample size increases, so does the probability (with a given confidence level) that its mean is the true mean. Since the sample size associated with Avrum is larger than that associated with Moishe, applying the test to their respective samples entails that the probability that Avrum's sample mean deviates from the true mean (his character) is lower than the corresponding probability in Moishe's case. Thus, even though their respective samples are identical in both mean and standard deviation, the moral difference between Avrum and Moishe may be expressed.

15. Objection: Total Rightness

Related to the previous objection are two worries raised by Professor Rohit Parikh. I shall discuss the first in this Section, and the second in the next.

Conjunction is defined as a measure of global moral ability and expressed by the *average* value of acts. But there are good reasons for thinking that it should be defined as a measure of total goodness and expressed by the *arithmetical sum* of moral values produced by an agent over a reasonably long time stretch. To see why, consider Hinda

and Golda. They are students of equal scholastic aptitude whose respective course work does and doesn't entitle them to a degree (Hinda has simply enrolled in college earlier than Golda). Since we are sometimes interested in whether a person is entitled to a degree, a measure of her total achievement is important.

Moreover, assigning a conjunctive act the arithmetical sum of the moral values of its conjuncts is (at least) as natural as assigning it the average of these values, since both assignments constitute an intuitive formal representation of revealing one's moral character. Why then prefer the average over the arithmetical sum of values (if conjunction is a function, it may not be assigned both values)? Or why not define a distinct operation (conjunction*) that yields as an output a measure of total rightness?

To begin with note that in order to find out whether a person is entitled to a degree, we compute the arithmetical sum not of her *grades*, but of the *credits* she has earned. If and only if the sum of credits meets a certain pre-determined criterion (say, 120 credits), the student is entitled to a degree. Now there appears to be no moral unit analogous to academic credits (moral values of acts correspond to grades). However, it wouldn't be absurd to ground the criterion for academic degrees not in credits, but in grades instead. In that case, students would be entitled to a degree upon reaching (say) 3,000 points. (If degree criteria were thus anchored in grades, slower students would end up taking more courses than strong ones.)

Let's then suppose that degree criteria are defined in terms of grades. Then the proposal under consideration is that it's important to set analogous rightness criteria. For such criteria would allow us to know whether a person is entitled to be called 'virtuous' or 'vicious.'

Well, in the first place, conjunctive* acts require an open-ended interval. As I argued in the previous Section, there is a powerful consideration against such an interval. Secondly, there is no reason why rightness criteria should be grounded in total rightness and not in average values of acts. Why not call a person 'virtuous' (or 'vicious') upon reaching not a pre-determined total rightness, but rather a certain average value?

16. Objection: Deflating Averages

A personal favorite is the second objection raised by Professor Parikh. Suppose Frieda saves a drowning child at risk of her life, which is very right. Later she returns some books to the library, which is mildly right. By returning the books she has brought down her average: *increasing rightness reduced* her moral ability. Or suppose that Frieda and Golda save a man from drowning. Later they return two and ten books to the library, respectively. It follows that because Golda produced *more* rightness than Frieda, her moral ability is *lower* than Frieda's. (To see this more clearly, suppose that saving a man from drowning is of value .7, whereas returning a book to the library is of value .1. Then Frieda's moral ability equals $1/3 (.7 + .1 \times 2) = .3$, while Golda's moral ability equals $1/11 (.7 + .1 \times 10) = .15$.) In short, my view on conjunction implies that under certain circumstances agents will reduce their overall moral ability by increasing rightness. But this is counter-intuitive; for presumably, agents can only increase their moral ability by increasing rightness.

This (admittedly neat) objection would be devastating if it meant that in order to maintain a high average, agents should refrain from performing mildly right actions after having performed very right ones. But it doesn't. This is because refraining from a mildly right action is wrong and thus of a *negative* value; it therefore brings down the

average even more than performing a mildly right action. To maintain a very high average, a person must encounter no obligations after having performed a highly right act—that is, he must be dead.

The second example calls for further treatment, for it means that Frieda is a better person than Golda even though it's the other way around (because Golda has produced more rightness). Well, it's clear I believe that the present difficulty arises from insufficient evidence. If Golda is indeed better than Frieda—if she consistently produces more rightness than her—their respective measures of moral ability would ultimately reflect this. When appropriately implemented then, measures of moral ability are unlikely to misrepresent the facts.

17. $\neg(\neg A \vee \neg B)$ or $\neg(\neg B \vee \neg A)$

We are ready to close a circle; we are now in a position to answer the question with which we opened the present Chapter: namely, is conjunction definable in terms of disjunction and negation?

As noted at the outset of this Chapter, in both two-valued logic and in the Łukasiewicz family of many-valued logic, $A \wedge B$ is equivalent to both $\neg(\neg A \vee \neg B)$ and $\neg(\neg B \vee \neg A)$ (because in the logics in question \vee commutes). Is $A \wedge B$ in the present system equivalent to either $\neg(\neg A \vee \neg B)$ or $\neg(\neg B \vee \neg A)$? (Recall that unless $v[A] = v[B]$, $v[\neg(\neg A \vee \neg B)] \neq v[\neg(\neg B \vee \neg A)]$; therefore, unless $v[A] = v[B]$, the question whether $A \wedge B$ is equivalent to $\neg(\neg A \vee \neg B)$ is distinct from the question whether it's equivalent to $\neg(\neg B \vee \neg A)$.)

Performing $A \vee B$ presumably amounts to committing A and refraining from B .³³

Analogously then, performing $\neg(\neg A \vee \neg B)$ amounts to refraining from refraining-from- A -and-refraining-from-refraining-from- B . Therefore, $A \wedge B$ is equivalent to $\neg(\neg A \vee \neg B)$ iff committing A and committing B means refraining from refraining-from- A -and-refraining-from-refraining-from- B . Does committing A and committing B mean refraining from refraining-from- A -and-refraining-from-refraining-from- B ? Since I don't understand what 'refraining-from-refraining-from- B ' means ($\neg\neg B$ is not a wff), the question makes no sense. For the same reasons, it makes no sense to ask whether $A \wedge B$ is equivalent to $\neg(\neg B \vee \neg A)$; I take this to be obvious.

The same conclusion can be arrived at via a different route. $v[A \wedge B] = \frac{1}{2} (v[A] + v[B])$. Does $v[\neg(\neg A \vee \neg B)]$ too equal $\frac{1}{2} (v[A] + v[B])$? Does $v[\neg(\neg B \vee \neg A)]$? You will recall that

If $v[A] + v[\neg B] > v[B] + v[\neg A]$, then

- (i) $v[A \vee B] = v[A]$
- (ii) $v[B \vee A] = v[B] + v[\neg A]$

If $v[A] + v[\neg B] < v[B] + v[\neg A]$, then

- (i) $v[A \vee B] = v[A] + v[\neg B]$
- (ii) $v[B \vee A] = v[B]$

The next step then is finding out whether $v[A] + v[\neg B] > v[B] + v[\neg A]$ or vice versa.

Since in the present case A and B are $\neg A$ and $\neg B$ (respectively), we in effect would like to find out whether $v[\neg A] + v[\neg\neg B] > v[\neg B] + v[\neg\neg A]$ or vice versa. To determine that, we need to compute the values of $\neg\neg A$ and $\neg\neg B$. Yet $\neg\neg A$ and $\neg\neg B$ are not wffs of the

³³ See also Chapter 3, Section 8.

present framework. Hence, it makes no sense to ask whether $v[\neg(\neg A \vee \neg B)]$ or $v[\neg(\neg B \vee \neg A)]$ equal $\frac{1}{2}(v[A] + v[B])$.

The question wouldn't make sense even if we stipulate not what $v[A]$ and $v[B]$ are, but rather what $v[\neg A]$ and $v[\neg B]$ are. To see this, suppose that $v[\neg A] = -.51$ and $v[\neg B] = -.68$. Then $v[A] = .3$ or $.7$, while $v[B] = .4$ or $.6$. To determine whether $v[A] + v[\neg B] > v[B] + v[\neg A]$ or vice versa, we must consider all four combinations of $v[\neg A]$, $v[\neg B]$, $v[A]$, and $v[B]$.³⁴

Consider first the case where $v[A] = .3$ and $v[B] = .4$. Then $v[A] + v[\neg B] = .3 - .68 = -.38$ whereas $v[B] + v[\neg A] = .4 - .51 = -.11$. This means that $v[A] + v[\neg B] < v[B] + v[\neg A]$. Or that $v[\neg A] + v[B] > v[\neg B] + v[A]$. Therefore, $v[\neg A \vee \neg B] = v[\neg A]$, which means that $v[\neg(\neg A \vee \neg B)] = v[\neg\neg A]$. Therefore again, $v[\neg B \vee \neg A]$ —which we must compute in order to determine $v[\neg(\neg B \vee \neg A)]$ —equals $v[\neg B] + v[\neg\neg A]$. But $\neg\neg A$ is not a wff of the present framework.

Next consider the case where $v[A] = .7$ and $v[B] = .4$. Then $v[A] + v[\neg B] = .7 - .68 = .02$ whereas $v[B] + v[\neg A] = .4 - .51 = -.11$. This means that $v[A] + v[\neg B] > v[B] + v[\neg A]$. Or that $v[\neg A] + v[B] < v[\neg B] + v[A]$. Therefore, $v[\neg A \vee \neg B]$ —which we must compute in order to determine $v[\neg(\neg A \vee \neg B)]$ —equals $v[\neg A] + v[\neg\neg B]$. Therefore again, $v[\neg B \vee \neg A] = v[\neg B]$, which means that $v[\neg(\neg B \vee \neg A)] = v[\neg\neg B]$. But $\neg\neg B$ is not a wff of the present system.

³⁴ Of course, $v[\neg A]$ and $v[\neg B]$ remain the same throughout; only $v[A]$ and $v[B]$ change.

Since these two cases exhaust the four possibilities offered by the relevant segment of the evaluation procedure for \mathbf{V} , there is no need to consider the remaining combinations of $v[\neg A]$, $v[\neg B]$, $v[A]$, and $v[B]$.

It makes no sense then to ask whether $A \wedge B$ is equivalent to $\neg(\neg A \mathbf{V} \neg B)$ or $\neg(\neg B \mathbf{V} \neg A)$. Yet conjunction and disjunction may be inter-definable in a different manner, a possibility that I turn to explore next.

18. $A \wedge \neg B$

As noted twice already,³⁵ since performing $A \mathbf{V} B$ presumably amounts to committing A and refraining from B , it stands to reason that $A \mathbf{V} B$ is equivalent to $A \wedge \neg B$. Is this indeed the case?

$v[A \wedge \neg B] = \frac{1}{2} (v[A] + v[\neg B])$. $v[A \mathbf{V} B]$ on the other hand equals (1) $v[A]$ when $v[A] + v[\neg B] > v[B] + v[\neg A]$, and (2) $v[A] + v[\neg B]$ when $v[A] + v[\neg B] < v[B] + v[\neg A]$. In neither case then does $v[A \mathbf{V} B]$ equal $v[A \wedge \neg B]$; which means that $(A \mathbf{V} B) \Leftrightarrow (A \wedge \neg B)$ fails in the present system.

One might wonder: since performing $A \mathbf{V} B$ amounts to committing A and refraining from B , why not *stipulate* that $v[A \mathbf{V} B]$ equals $\frac{1}{2} (v[A] + v[\neg B])$? In accordance with the valuation procedure for ‘or’ as well as the previous paragraph, I shall divide my response into two parts. First of all, if $v[A \mathbf{V} B]$ equals $\frac{1}{2} (v[A] + v[\neg B])$ when $v[A] + v[\neg B] > v[B] + v[\neg A]$, then $v[\neg B]$ affects the value of $A \mathbf{V} B$. This result conflicts with the requirement that agents should not be punished for appropriately resolving dilemmas which (I am assuming) they involuntarily encounter.³⁶ Giving up the

³⁵ In Chapter 3, Section 8, and Section 17 above.

³⁶ See Chapter 3, Section 6.

requirement in order to secure that $v[A \vee B]$ always equals $v[A \wedge \neg B]$ seems to exact too high a price from our moral intuitions. (On the other hand, holding on to the requirement entails that $v[A \vee B] \neq v[A \wedge \neg B]$ even though performing $A \vee B$ amounts to committing A and refraining from B . Isn't *that* too costly a concession? I think not. Recall that performing $A \vee B$ amounts to committing A and refraining from B because committing A entails refraining from B : that is, agents are *forced* to refrain from B . The valuation function for $A \wedge \neg B$ on the other hand presupposes that agents can perform both A and B : that they refrain from B by *choice*. Why shouldn't the difference between $A \vee B$ and $A \wedge \neg B$ translate into a discrepancy between $v[A \vee B]$ and $v[A \wedge \neg B]$?)

Secondly, if $v[A \vee B]$ equals $\frac{1}{2}(v[A] + v[\neg B])$ when $v[A] + v[\neg B] < v[B] + v[\neg A]$, it follows that the degree of wrongness assigned to some failures to maximize net moral value is arbitrarily cut by half. I submit that *this* result may seem a *reasonable* price to pay for securing that $v[A \vee B]$ equals $v[A \wedge \neg B]$ when $v[A] + v[\neg B] < v[B] + v[\neg A]$.

Should we then hold that $v[A \vee B]$ equals $v[A \wedge \neg B]$ whenever agents fail to maximize net moral value? Two cases remain to be considered: (1) $v[B \vee A]$ when $v[A] + v[\neg B] < v[B] + v[\neg A]$, and (2) $v[\vee A, B]$ when either $v[\neg A] > v[\neg B]$ or $v[\neg B] > v[\neg A]$. (1) is entirely analogous to the case I discussed in the previous paragraph: hence, whatever I said of the latter equally applies (*mutatis mutandis*) to the former.

Concerning (2), you will recall that $\vee A, B$ reads: refraining from both A and B . It therefore amounts to refraining from A and refraining from B . Thus, for $v[\vee A, B]$ to equal $v[\neg A \wedge \neg B]$, it must equal $\frac{1}{2}(v[\neg A] + v[\neg B])$. And it does.³⁷

³⁷ See Chapter 3, Section 7.

Some might feel that in that case too, securing that $v[\mathbf{V} A, B] = v[\neg A \wedge \neg B]$ exacts too high a cost from our moral intuitions. To see why, suppose that Golda refrains from the duty to save a life at a negligible sacrifice (A) and the duty to keep a trivial promise (B), where A and B are conflicting duties. Now $v[\neg A]$ is very low while $v[\neg B]$ is only slightly low. Therefore, if $v[\mathbf{V} A, B]$ is affected by $v[\neg A]$ alone, it too would be very low. On the other hand, if $v[\mathbf{V} A, B]$ is affected by both $v[\neg A]$ and $v[\neg B]$ —as it is if it equals $\frac{1}{2} (v[\neg A] + v[\neg B])$ —it wouldn't be as low. To take a numerical example, suppose that $v[\neg A] = -.85$ and $v[\neg B] = -.2$. If $v[\mathbf{V} A, B] = v[\neg A]$, it equals $-.85$. If however $v[\mathbf{V} A, B] = \frac{1}{2} (v[\neg A] + v[\neg B])$, it equals $-.525$. It turns out then that if $v[\mathbf{V} A, B] = \frac{1}{2} (v[\neg A] + v[\neg B])$, Golda benefits from refraining from keeping a trivial promise. As I say, in light of this consequence some may deny that $v[\mathbf{V} A, B]$ equals $v[\neg A \wedge \neg B]$ (and perhaps require that the value assigned to it be the lowest possible within the constraints imposed by—on the one hand—the system's interval and—on the other hand— $v[\neg A]$ and $v[\neg B]$). Others may be willing to bite the bullet.

The upshot of the foregoing discussion is that *provided one is willing to compromise certain moral intuitions*, $(A \mathbf{V} B) \Leftrightarrow (A \wedge \neg B)$ holds whenever agents fail to maximize moral value. Or—less sweepingly—that $(A \mathbf{V} B) \Leftrightarrow (A \wedge \neg B)$ holds whenever one is willing to compromise certain moral intuitions.

19. Summary

Let's now turn to summarize this Chapter. A conjunctive act consists in performing at least two distinct moral actions either concurrently or at different times. It amounts to revealing or exercising one's moral character, consists of no bodily movements over and above the ones required to carry out the conjuncts, and is an entity

though composed of several distinct actions.

Revealing or exercising one's moral character—a conjunctive action—is assigned the value of a person's moral character (which is her overall or global moral ability): that is, the weighted average of the values of the conjuncts. Thus, the truth-table for conjunction is:

$$v[A_1 \wedge A_2 \wedge \dots \wedge A_n] = n_i \bar{x}_i / (n_i + n_j) + n_j \bar{x}_j / (n_i + n_j)$$

Chapter 5: 'By'

1. Introductory Remarks

The sheriff's saving hundreds of lives by killing an innocent man surely differs morally from his saving hundreds of lives by donating his entire fortune. Rebecca's flattering Jacob by insulting Esau clearly differs from her flattering Jacob by praising his wit. And the government's maintaining law and order by oppressing political dissidents differs from its maintaining law and order by securing the well-being of all members of society. In short, the moral nature of means for praiseworthy ends matters.

Of course, the moral nature of means for *despicable* ends matters too. Insulting Esau by criticizing his judgment certainly differs from insulting Esau by flattering Jacob. Morally speaking then, 'by' gives rise to four types of cases. They are represented by the following table.

	Means	Good	Bad
End			
Good			a b
Bad		c	d

Fig. 5.1. 'By'

2. Unifiers and Multipliers

These cases constitute a *prima facie* case for considering 'by' a major moral connective worthy of formal treatment. It's a matter of course that taking even a first stab at this task requires clarity concerning the nature of 'by.'

Unifiers argue that 'by' signifies identity, thus obtaining only between actions and themselves.¹ Multipliers on the other hand believe that 'by' combines numerically distinct though intimately related actions.²

Sorting out the dispute between unifiers and multipliers then is a prerequisite for further inquiries into 'by.' As it turns out, this process occupies the bulk of this Chapter.

3. Multipliers or Goldman's Theory of Human Action

In his *Theory of Human Action* Alvin Goldman maintains that actions are instantiations of act-properties by an agent at a time.³ Two act-tokens are therefore identical iff they involve the same agent, the same property, and the same time. Since insulting Esau is a property distinct from flattering Jacob, Rebecca's insulting of Esau is distinct from her flattering of Jacob.

Goldman suggests that 'by' signifies a relation he calls 'generation.'⁴ It holds between ordered pairs of act-tokens of the same agent, such as Rebecca's insulting of Esau and her flattering of Jacob, John's extending his arm out the car window and John's signaling a turn, and so on.

If a pair of act-tokens A and A' are related by generation, three temporal features characterize its members. The first is that neither act is subsequent to the other: it's incorrect to say that the agent did one *and then* (or *and later*) did the other. The second is that the acts are not co-temporal: it's incorrect to say that the agent did one of the acts

¹ 'By' need not hold between actions only. Its relata may also be events and facts. However, only actions (*moral* actions, to be precise) are of interest for present purposes.

² The titles 'unifiers' and 'multipliers' are due to Irving Thalberg ("Singling Out Actions, their Properties and Components" and "When Do Causes Take Effect").

³ This view is also held by Jaegwon Kim ("Events as Property Exemplification").

⁴ The 'by' locution always signifies generation, but not all instances of generation can be naturally expressed by it. More on this matter shortly.

while also doing the other.⁵ Finally, the acts are performed during the same time interval: that is, neither is a temporal part of the other. For example, playing the C-scale is composed of a series of shorter subsequent acts (playing note C followed by playing note D, and so on). Each of these acts is a temporal part of playing the C-scale. Therefore, none of them is generationally related to the longer act of playing the C-scale.⁶

Goldman distinguishes four kinds of generation: causal, simple, conventional, and augmentative. Causal generation holds where an agent's act-token *A* has an effect *E* thanks to which she may be credited with performing act *A'*. For example, Cohen's flipping the switch has the effect of the light going on, in virtue of which he may be credited with the act of turning the light on.⁷ In general, act-token *A* of agent *S* causally generates act-token *A'* of his only if *A* causes *E* and *A'* consists in *S*'s causing *E*.⁸

Note that *A* causes *E* and causally generates *A'*; that is, *A* does *not* cause *A'*. Thus, Levi's flipping the switch (*A*) causes the light to go on (*E*), but *not* his turning the light on (*A'*); the latter is caused (say) by his desire to alert a prowler to the fact that he is home.⁹ Also note that the notion of causal generation includes cases where an action is merely a contributing factor of an effect: if Esther contributed to the panic by whistling, her whistling causally generated the panic.¹⁰

Conventional generation occurs where, thanks to conventions (rules, norms, social practices), an act *A'* can be ascribed to an agent *S* upon his performing another act *A*.

⁵ See Chapter 4, fn. 4.

⁶ *A Theory of Human Action*, 21-2.

⁷ *Ibid.*, 10-11.

⁸ *Ibid.*, 22-3.

⁹ *Ibid.*, 23-4. This means that Goldman is in disagreement with Arthur Danto. For Danto maintains that if I close the door by moving my hand, my nonbasic action of closing the door is caused by my (numerically distinct) basic action of moving my hand ("What We Can Do" and "Basic Actions").

¹⁰ *A Theory of Human Action*, 24.

Thus, *S*'s moving his queen to king-knight-seven conventionally generates *S*'s checkmating his opponent. An act-token *A* then conventionally generates act-token *A'* only if the performance of *A*, together with a convention saying that *A* counts as *A'*, guarantees *A'*.¹¹

Simple generation involves neither conventions nor causal connections between *A* and some effect. It occurs when certain circumstances *C* together with *A* suffice for *A'*. Thus, if *S* jumps seven feet (*A*) when George has jumped six feet (*C*), *S* outjumps George (*A'*). Likewise, if *S*—believing *p*—says not-*p*, *S* is lying. Note that the circumstances of the latter example consist in the agent's being in a certain mental state.¹²

The following pairs are related by augmentative generation: *S*'s saying 'hello' and *S*'s saying 'hello' loudly; *S*'s running and *S*'s running at 8 m.p.h.; *S*'s extending his arm and *S*'s extending his arm through the window. Since the 'by' locution is inapplicable to augmentation generation (*S* does not run at 8 m.p.h. by running), 'by' doesn't refer to it. (This is the reason for fn. 4).¹³

As noted, if 'by' signifies a relation such as generation, it obtains between *numerically distinct* actions. However, compelling considerations suggest that 'by' signifies a relation that may hold only between things and themselves: namely, identity.

4. Unifiers

According to unifiers (or identity theorists; the latter is Goldman's title¹⁴), insulting Esau simply *is* flattering Jacob. The same applies, to alter the example, to 'Borg signified his joy by falling to his knees': falling to one's knees *is*—under the

¹¹ *Ibid.*, 25-6.

¹² *Ibid.*, 26-7. Jonathan Bennett argues that there is no need for conventional generation: Goldman's first three categories can be reduced to two—the causal and the noncausal (*Events and their Names*, 205-7).

circumstances (which I imagine were a triumph in a tennis tournament)—signifying joy.¹⁵ Again, if a man is replenishing the water supply of a house by operating a pump, his operating of the pump *is* his replenishing of the water supply.¹⁶ In all three cases therefore an agent is performing exactly one act, only it is considered *under different descriptions*. In Anscombe's own words (they are simply too colorful to be left out),

A man is pumping water into the cistern which supplies the drinking water of [an inhabited] house. Someone has found a way of systematically contaminating the source with a deadly cumulative poison whose effects are unnoticeable until they can no longer be cured. ...[H]e has revealed [this] to the man who is pumping. ...

This man's arm is going up and down, up and down. Certain muscles, with Latin names which doctors know, are contracting and relaxing. ... The moving arm is casting a shadow on a rockery where at one place and from one position it produces a curious effect as if a face were looking out of the rockery. Further, the pump makes a series of clicking noises, which are in fact beating out a noticeable rhythm. ...

Are we to say that the man who (intentionally) moves his arm, operates the pump, replenishes the water supply, poisons the inhabitants, is performing *four* actions? Or only one? ... moving his arm up and down with his fingers round the pump *is*, in these circumstances, operating the pump; and in these circumstances, it *is* replenishing the house water supply; and, in these circumstances, it *is* poisoning the household.

So there is one action with four different descriptions... each related to the next as description of means to end...¹⁷

Anscombe, who authored the identity thesis, was soon joined by Donald Davidson.¹⁸ He writes:

¹³ *A Theory of Human Action*, 28-9.

¹⁴ *Ibid.*, 2.

¹⁵ The example is Jonathan Bennett's (*Events and their Names*, 188.)

¹⁶ The example is of course G. E. M. Anscombe's (*Intention*, 45).

¹⁷ *Intention*, 37, 45-6.

I flip the switch, turn on the light, and illuminate the room. Unbeknownst to me I also alert a prowler to the fact that I am home. Here I need not have done four things, but only one, of which four descriptions have been given.¹⁹

And also:

But what is the relation between my pointing the gun and pulling the trigger, and my shooting the victim? The natural and, I think, correct answer is that the relation is that of identity.²⁰

Apart from its *prima facie* plausibility, the identity thesis has two considerations to recommend it. The first is that the thesis makes sense of excuses. ‘I didn’t know it was loaded’ means that since I didn’t know the gun was loaded, I didn’t know that my excusable pulling of the trigger would be (identical to) my deplorable shooting of the victim. Likewise, ‘I did it in self-defense’ means that I knew or believed that my deplorable shooting of the victim would be (identical to) my excusable shooting of her in self-defense.²¹

Trouble: these excuses make sense also when the actions involved are taken to be numerically distinct, for they could read: ‘I didn’t know that if I did one of these actions I would inevitably do the *other*.’²²

The second consideration (also Davidson’s) came to be known as the ‘no further effort argument.’ It says that if the queen killed the king by pouring poison in his ear, and we deny that the poison pouring is the killing, we would have to admit that—after pouring poison in the king’s ear—there remains some deed for the queen to do: but this is

¹⁸ In addition to Anscombe and Davidson, proponents of the identity thesis include Joel Feinberg (“Action and Responsibility,” 146). Feinberg however also says things that seem to contradict the identity thesis, especially in connection with his discussion of ascriptions of simple agency (*ibid.*, 147).

¹⁹ “Actions, Reasons and Causes,” 4.

²⁰ “The Logical Form of Action Sentences,” 109.

²¹ Davidson *ibid.*, 109; “The Individuation of Events,” 165.

²² This objection of Lawrence Davis is mentioned in Bennett, *Events and their Names*, 193.

absurd: the queen has done her work; it only remains for the poison to do its.²³ In slightly other words, denying the identity thesis implies that there is something left for the queen to do when in fact there is no such thing.

‘The queen moved her hand in such a way as to pour poison in the king’s ear’ and ‘she did something that caused the death of—i.e., killed—the king’ are nothing but two descriptions of the same action, which is the moving of her hand by the queen.²⁴ Both descriptions include reference to consequences of the action—the former to the poison pouring, the latter to the killing of the king—and we may extend responsibility or liability for an action to responsibility or liability for its consequences. But the consequences are not themselves actions.

We must conclude, perhaps with a shock of surprise, that our primitive actions, the ones we do not do by doing something else, mere movements of the body—these are all the actions that there are. We never do more than move our bodies: the rest is up to nature.²⁵

We must also conclude that the language we use to describe actions exhibits a feature that Joel Feinberg has dubbed ‘the accordion effect’²⁶: like the folding musical instrument, a person’s act may be squeezed down to a minimum (primitive action) or else stretched out to include mention of effects—that is, it may be described as narrowly or as broadly as we please.

There are, in fact, a great many tunes we can play on the accordion. We could start with, ‘The queen moved her hand’ and pull to the right by adding, ‘thus causing the vial to empty into the king’s ear’; and now another tug, ‘thus causing the poison to enter the body of the king; and finally

²³ “Agency,” 57-8.

²⁴ *Ibid.*, 58.

²⁵ *Ibid.*, 59. Davidson’s primitive actions are Feinberg’s simple (“Action and Responsibility,” 145) and Arthur Danto’s basic (“What We Can Do” and “Basic Actions”).

²⁶ “Action and Responsibility,” 145.

(if we have had enough—for the possibilities for expression are without limit), ‘thus causing the king to die’. This expression can be shortened in many ways, into the center, the left, or the right components, or any combination. For some examples: ‘The queen moved her hand thus causing the death of the king’ (the two ends); or, ‘The queen killed the king (collapse to the right); or, ‘The queen emptied the vial into the king’s ear’ (the center). There is another way to pull the instrument out, too: we could *start* with, ‘The queen killed the king’, adding ‘by pouring poison in his ear’, and so on—addition to the left.²⁷

5. Against Unifying (1): Symmetry and Reflexivity

Goldman has done much (probably more than anyone else) to prove the identity thesis wrong (because it is inconsistent with his own method of action individuation). His three objections to the thesis are the topic of Sections 5-7. Two further objections are the subject matter of Sections 8-9.

Goldman’s first objection is that the identity relation is symmetric (if *a* is identical to *b*, *b* is identical to *a*) and reflexive (if *a* is identical to *b*, *a* is identical to itself), whereas the ‘by’ relation is asymmetric and irreflexive. The queen killed the king by poisoning him, but she did *not* poison him by killing him—this shows that the ‘by’ relation is asymmetric. The queen killed the king by poisoning him, but she did *not* poison him by poisoning him—this shows that the ‘by’ relation is irreflexive.²⁸

In her “Individuating Actions” Judith Jarvis Thomson replies that this objection rules out Goldman’s fine-grained procedure of individuating actions, a procedure for

²⁷ Davidson, “Agency,” 58-9. As Davidson notes, some authors have pointed out that it is not *always* possible to replace ‘*a* did *X* and thereby caused *Y*’ by ‘*a* did *Y*.’ For example, John Atwell writes that ‘Jenkins caused Murphy to be promoted by informing the corporation head of Murphy’s superior performance record’ cannot be replaced by ‘Jenkins promoted Murphy’ if Jenkins doesn’t have the authority to make promotions (“The Accordion Effect Thesis,” 37).

²⁸ *A Theory of Human Action*, 5.

which it was designed to pave the way. In particular, the identity theorist holds that if Sebastian is replenishing the water supply of a house by operating a pump, then

(1) Sebastian's operating of the pump is his replenishing of the water supply

In response Goldman argues that while

(2) Sebastian replenished the water supply by operating the pump

is true.

(3) Sebastian operated the pump by replenishing the water supply

is false (this is why the 'by' relation is asymmetric), and

(4) Sebastian operated the pump by operating the pump

is simply odd (this is why the 'by' relation is irreflexive).

Goldman presumably thinks that (1) should license substitution into (2) to yield (3) and (4). And Jarvis Thomson asks: How could it? (2) doesn't contain either of the singular terms 'his operating of the pump' or 'his replenishing of the water supply,' so what is there to substitute?²⁹

Perhaps (Jarvis Thomson continues) Goldman thinks that (2) is analyzable into

(2') Sebastian's replenishing of the water supply has the 'by' relation to his operating of the pump

Could (2') be a paraphrase of (2)? Suppose that Sebastian has been pumping away every morning for weeks, and only today, for the first time, did he replenish the water supply by operating the pump. Then (2) is true. But if in saying (2') one refers to yesterday's pumping, (2') is false. Hence, (2') could *not* be a paraphrase of (2).³⁰

²⁹ P. 775. Anscombe, similarly, points out that the question whether two phrases are true of one and the same action cannot be settled by pointing to their non-substitutivity if they are anyway not singular terms ("Under a Description," 224).

³⁰ "Individuating Actions," 775-6.

Perhaps then Goldman thinks that (2) is analyzable into

(2'') There is an x and there is a y such that x is a replenishing of the water supply by Sebastian, and y is an operating of the pump by Sebastian, and x has the 'by' relation to y .³¹

Well, while (2'') seems a kosher paraphrase of (2), *Goldman* cannot agree that it is.

Specifically, (2'') is a Davidson-type construal of (2). You will recall that according to

Davidson,

(5) Jones buttered the toast

is analyzed as follows

(5') $(\exists a)$ (Buttering (Jones, the toast, a))

and

(6) Jones buttered the toast with a knife

is analyzed as follows

(6') $(\exists a)$ {Buttering (Jones, the toast, a) \wedge $(\exists b)$ [knife (b) \wedge with (a , b)]}³²

This analysis embodies the notion that buttering of the toast by Jones is identical to buttering of the toast with a knife by Jones. So if (2) is analyzed as (2''), replenishing of the water supply by Sebastian is identical to replenishing of the water supply by operating the pump by Sebastian.

But Goldman denies this identity. On his view, two actions are identical only if they exemplify the same property. Thus, replenishing the water supply by Sebastian is *distinct* from replenishing of the water supply by operating the pump by Sebastian (because the act properties '...is replenishing the water supply' and '...is replenishing the

³¹ *Ibid.*, 776. I omit the phrase 'and x is before NOW' which Jarvis Thomson has tacked onto (2'').

³² See Chapter 4, Section 6.

water supply by operating the pump' are different).³³ Hence, if—as Goldman's first objection requires—(2) is analyzed as (2''), his fine-grained method of individuating actions is ruled out.

6. Against Unifying (2): Causal Properties of Acts

It seems then that Goldman's first objection need not worry the unifier. Need his second objection worry her? Goldman's second objection—a *line* of objections as a matter of fact—relies on the assumption (which Davidson shares explicitly³⁴ and Anscombe implicitly³⁵) that identical actions have all their causal properties in common. If by playing the piano Jones both puts Smith to sleep and awakens Brown, then—according to the identity thesis—Jones's playing the piano is identical to his putting Smith to sleep and also to his awakening Brown. But Jones's playing the piano has an effect that his putting Smith to sleep does not, and that is awakening Brown. This means that 'Jones's playing the piano' and 'Jones putting Smith to sleep' cannot be two descriptions of one and the same action, and must therefore refer to numerically distinct actions.³⁶

Davidson would in all likelihood respond that there is but a single action in this case. It is the moving of his fingers by Jones, which *resulted* in playing the piano, which in turn resulted in putting Smith to sleep and in awakening Brown. The moving of his fingers by Jones is identical to—and hence has the same effects as—the moving of his fingers by Jones which resulted in playing the piano (which resulted in putting Smith to sleep and in awakening Brown). Again, if by moving his hand Cohen is both frightening

³³ Jarvis Thomson. "Individuating Actions," 779.

³⁴ "The Individuation of Events," 179.

³⁵ "Under a Description," 226.

³⁶ *A Theory of Human Action*, 3–4.

away a fly and waving goodbye, the moving of his hand by Cohen is identical to—and thus has the same effects as—both the moving of his hand that resulted in frightening away a fly and the moving of his hand that resulted in waving goodbye.

Here is another instance of Goldman's second objection. If Jones killed Smith by pulling the trigger, then—according to the identity thesis—his pulling of the trigger is identical to his killing of the victim; but this can't be so, for while the trigger pulling caused the event of the gun's going off, the killing did not.

Anscombe replies that 'The killing caused the gun to go off' simply means that the act which (as things turned out) was the killing caused the gun to go off. This case is analogous to 'The widow stuck her husband with a knife' which simply means that the lady who (as things turned out) is now a widow stuck her husband with a knife.³⁷

It has been retorted that Anscombe in effect argues that the trigger pulling is (as things turned out) the same as the killing, and that since the relation '...is (as things turned out) the same as...' must be that of identity, her reply to Goldman is circular.³⁸ Be that as it may, the identity theorist can reply to Goldman *à la* Davidson, pointing out that the act—i.e., hand movement—which resulted in the killing is the same as that which resulted in the trigger pulling; therefore—like the latter—it resulted in the gun's going off.

7. Against Unifying (3): The Tense Problem

It appears then that those objections of Goldman's, which rely on the assumption that identical actions have all their causal properties in common and which focus on *effects*, don't prove the identity thesis wrong. Perhaps the objections that focus on *causes*

³⁷ "Under a Description," 226-7.

³⁸ For example, Roberto Casati and Achille Varzi, *Events*, 18.

would. Suppose that Cohen screwed a bulb into the socket and shortly afterwards Levi turned the light on by flipping the switch. According to the identity thesis, Levi's turning the light on is identical to his flipping the switch (that is, to the moving of his hand which resulted in flipping the switch). But then, the Levi hand movement which resulted in flipping the switch should have the same causes as the movement which resulted in turning the light on. This however isn't the case, for Cohen's screwing a bulb into the socket played no causal role in—wasn't a necessary condition for—Levi's flipping the switch, but *was* a necessary condition for his turning the light on. We must accordingly conclude that Levi's flipping the switch is not identical to his turning the light on.

With respect to this objection too Davidson would most probably point out that the hand movement which resulted in flipping the switch is identical to the hand movement which resulted in turning the light on (and therefore flipping the switch and turning the light on are identical actions). But how would he explain the fact that a functional bulb is necessary for turning the light on but not for flipping the switch? How—in other words—would he account for the fact that two different sets of causal conditions are necessary for flipping the switch and for turning the light on?

The objection under consideration is closely related to (presumably a generalization of) what Jonathan Bennett has called 'the *actio praecox* argument'³⁹ and Jarvis Thomson—who discovered the argument—'the tense problem.'⁴⁰ If Dirk repelled the May floodwaters by building a wall the previous September, then—according to the identity thesis—his repelling of the waters is identical to his construction of the wall: but the repelling of the waters cannot be, as the construction of the wall is, complete before

³⁹ *Events and their Names*, 195.

⁴⁰ "The Time of a Killing," 118.

the waters have risen.⁴¹ That is, floodwaters are a necessary condition for the repelling of the waters, but not for the construction of the wall. The same applies of course to the poison pouring and the killing: the king's death is necessary for the latter alone: thus, if the poisoning took place at 2:00 PM and the king died at 4:00 PM, then at 3:00 PM the poisoning may be said to have occurred, whereas the killing may not.⁴²

Davidson is unmoved. He writes:

To describe an event as a killing is to describe it as an event (here an action) that caused a death, and we are not apt to describe an action as one that caused a death until the death occurs; yet it may be such an action before the death occurs. (And as it becomes more certain that a death will result from an action, we feel less paradox in saying, 'You have killed him'.) [In a footnote]...the poisoned Hamlet, in killing the King, avenges, among other murders, his own. This he could not do if he had not already been murdered.⁴³

But the notion that Hamlet fought an entire sword-fight after being murdered has a bizarre ring to it (though, come to think of it, a man dies *after* the performance of any act that could ever be considered an act of killing him⁴⁴). Have identity theorists offered more palatable solutions to the tense problem?

Anscombe argues that the fact that at 3:00 PM the killing may not be said to have occurred certainly doesn't mean that *the act of killing* had not happened until the king was dead. It means only that although that act has occurred which, as things turn out, will prove to have been an act of killing, things have not yet turned out so. In other

⁴¹ *Events and their Names*, 195.

⁴² "The Time of a Killing," 116-9. Similar problems arise in analogous manner. The date problem is that while 'The queen poured poison in the king's ear at 2:00 PM' is true, 'The queen killed the king at 2:00 PM' is false. The temporal order problem is that 'The king's death occurred two hours after the queen poisoned him' is true, whereas 'The king's death occurred two hours after the queen killed him' is false (Jarvis Thomson, *ibid.*, 118-9).

⁴³ "The Individuation of Events," 177.

⁴⁴ Anscombe's point ("Under a Description," 228).

words, something further has to happen in order for the poisoning to *have been* an act of killing; but *pace* Jarvis Thomson, nothing further has to happen in order for the killing to *have taken place*.⁴⁵

Bennett, in a similar vein, suggests that the construction of the wall and the repelling of the waters are the same action: namely, the action of Dirk's constructing the wall, which later—indeed after it has ceased to exist—*became* his repelling of the waters.⁴⁶ Likewise, the poison pouring and the killing are the same action, that is, the poison pouring which later—with the king's death—became the killing. As Bennett is well aware, his position implies that an action may change its status after it has occurred. This notion, he maintains, may but need not raise eyebrows, for just as actions may become famous or forgotten long after they have ceased to exist, they may also become the cause of so-and-so when no longer taking place. Indeed, anything—and not just events and actions—can acquire a relational property after it has ceased to exist: Johanna Rosine Wagner had been dead for thirty-five years before she first became describable as 'mother of someone who has composed *Parsifal*.'⁴⁷

One might protest that a poison pouring which later became the cause of a death is no longer itself: its identity doesn't persist through the change. A bodily movement that later became the cause of an intentional death is murder: that is, an action which—due to its intended objective effect—has transformed into something *else*. Indeed, legal practice confirms this observation: at 3:00 PM the queen would have been charged with

⁴⁵ *Ibid.*, 227.

⁴⁶ *Events and their Names*, 197. Bennett first presented this account in his "Shooting, Killing and Dying." In the later *Events and their Names* he withdraws from it, endorsing an essentially Goldmanian approach to the analysis of the 'by'-locution (299, Chapter 14).

⁴⁷ *Ibid.*, 197-8.

attempted murder, at 4:00 PM with murder. This presumably means that Bennett's suggestion doesn't draw the sting out of the tense problem.

Well, not quite. An action's *status* (legal or other) may change due to its objective consequences (whether intentional or not). But an action may not thereby *multiply*. Once again, *one and the same* bodily movement amounts to both the poison pouring and the killing; it doesn't cause or generate *further* actions.

In response one might submit that this is so, and argue that if an action's status can change due to its consequences, so can its moral value. After all, a bodily movement intended to cause death presumably worsens morally once in fact bringing the death about. This means that even though 'by' never combines numerically distinct actions (because it expresses identity), the *moral value* of (say) 'killing by torturing' is a product of two (potentially different) moral values—one contributed by the killing, the other by the torturing. Again, though aiding the poor by stealing from the rich (Robin Hood) is a single act (or a single sequence of actions), its moral value is determined by two moral values, that of the aiding and that of the stealing.

Well, in the first place, this suggestion is at odds with my working assumption: namely, that objective consequences are morally irrelevant.⁴⁸ For this assumption implies that an action's *moral* status may *not* change due to its objective consequences: murder and attempted murder for example are morally alike. Secondly and more importantly, if a single act amounts to both aiding the poor and stealing from the rich, why not take it to be simply an action with two consequences—one good, the other bad—

⁴⁸ Possible misgivings concerning this assumption—worries involving moral luck and strict liability—are mentioned in Chapter 1, Section 6 as well as in Chapter 7, Section 7.

whose *single* moral value is affected by both consequences? I discuss this notion further in the Sections to follow.

8. Against Unifying (4): Non-extensional Temporal Functions

Two additional objections to the identity thesis remain to be considered. The first is Jarvis Thomson's. Jarvis Thomson argues that in order to eliminate the tense problem, the unifier must allow that 'the time of completion of x 's verbing y ' is a non-extensional expression. Specifically, the unifier can proceed by paraphrasing

- (1) The queen's poisoning of the king has occurred
- (2) The queen's killing of the king has occurred

as, respectively.

- (3) There is a time u and a time v such that u is the time of initiation of the queen's poisoning of the king and v the time of its completion, and u and v both precede 3:00 PM
- (4) There is a time u and a time v , such that u is the time of initiation of the queen's killing of the king and v the time of its completion, and u and v both precede 3:00 PM

Now you will recall that at 3:00 PM (1) and (3) are true, while (2) and (4) are false. Yet the unifier holds that

- (5) The poisoning is the killing

This means that the conjunction of (3) and (5) shouldn't entail (4). To achieve this result, the unifier must deny that 'the poisoning' and 'the killing' are inter-substitutable *salva veritate* in 'the time of completion of the poisoning (killing) of the king by the queen.'

She must, that is, allow that the latter expression is non-extensional. For then, (5) doesn't license substitution of 'the killing' for 'the poisoning' everywhere in (3); therefore, the conjunction of these statements doesn't entail (4). Given the tense problem then,

accepting the identity thesis requires conceding that ‘the time of completion of *x*’s verbing *y*’ is non-extensional.⁴⁹

As Jarvis Thomson herself recognizes however, this concession may not be all that major. For after all, there are a good many temporal functions which are—on any view—non-extensional. For instance, suppose that

(6) The father of the first set of quadruplets is the paternal grandfather of the first set of quintuplets.

(7) John became the father of the first set of quadruplets in 1890.

(6) and (7) do *not* entail

(8) John became the father of the first set of quintuplets in 1890.⁵⁰

Moreover, *pace* Jarvis Thomson, allowing that ‘the time of completion of *x*’s verbing *y*’ is a non-extensional expression is *not* the only solution for the tense problem available to the unifier. As the discussion in the previous Section indicates, other solutions are available to him.

9. Against Unifying (5): Moral Properties

The final objection concerns acts such as saving hundreds of lives by killing an innocent man as a scapegoat. Or helping the poor by stealing from the rich. According to the unifier, the killing is the saving (and the stealing is the helping). But saving hundreds of lives is right while killing an innocent bystander is wrong. Likewise, helping the poor is right whereas stealing from the rich is wrong. That is to say, if the identity theory is correct, the self-same act is both right and not right. Which is absurd. We must therefore deny that the saving is identical to the killing. Or that the helping is identical to

⁴⁹ “The Time of a Killing,” 128-9.

⁵⁰ *Ibid.*, 130.

the stealing.

Hardly a difficulty for a unifier. The killing and the saving are nothing but two consequences—the former bad, the latter good—of one and the same action. Ditto the helping and the stealing.

10. Truth-table

The foregoing discussion establishes that the unifier's position is superior to the multiplier's. 'By' then signifies identity, thus obtaining only between actions and themselves. Since an act may have but a single moral value, the truth-table for 'by' is:

$$v[d_1\{A\} \text{ by } d_2\{A\}] = v[(d_1 + d_2)\{A\}]$$

$d_1\{A\}$ is an action under a description. $v[d_1\{A\} \text{ by } d_2\{A\}]$ reads: the value of $\{A\}$ -under-a-description- d_1 by $\{A\}$ -under-a-description- d_2 . $v[(d_1 + d_2)\{A\}]$ reads: the value of $\{A\}$ -under-both- d_1 -and- d_2 . In English, the value of (say) helping the poor by stealing from the rich is the value of the (single) act that results in—or amounts to—both helping the poor and stealing from the rich.

11. $(d_1\{A\} \text{ by } d_2\{A\}) \text{ by } d_3\{A\}$

$d_1\{A\} \text{ by } d_2\{A\}$ is a wff. Are $(d_1\{A\} \text{ by } d_2\{A\}) \text{ by } d_3\{A\}$ and its like also wffs? In some moods I'm inclined to think that they aren't, for it appears to me that descriptions such as 'killing by shooting by pulling the trigger' or 'killing by shooting by pulling the trigger by moving one's finger' are hard or even impossible to parse.⁵¹ On other occasions however such descriptions seem almost crystal clear: which means that the symbolizations corresponding to them *are* wffs.

If they are wffs, the truth-table for 'by' is:

⁵¹ Compare with Chapter 2, Section 16. As noted there, the present view is reminiscent of Dorothy Edgington's discussion of some embedded conditionals.

$$v[(d_1\{A\} \text{ by } d_2\{A\}) \text{ by } \dots d_n\{A\}] = v[(d_1 + d_2 + \dots + d_n)\{A\}]$$

12. Moral Dilemmas Revisited: The Sheriff and the Innocent Bystander

Suppose that in a small, God forsaken American town a white woman has been raped and cruelly murdered. White residents of this town suspect that a black man committed the crime. Outraged and aching, they are about to take the law into their hands and lynch hundreds of black people. The local sheriff can prevent these killings by executing an innocent man as a scapegoat.⁵²

The sheriff then can prevent hundreds of deaths by killing an innocent man. Clearly, preventing hundreds of deaths is right, while killing an innocent wrong. Since the killing *is* the preventing, this means that there are reasons both for and against performing one and the same act. Since both committing and refraining from one and the same act is impossible, the scenario under consideration no doubt exhibits the defining features of a moral dilemma. Is it in fact a dilemma?

Before proceeding, I should like to point out that a great many other situations give rise to this question. Saving a mother's life by killing her fetus. Raising the suicide rate in the country by raising the level of education. Saving many lives by blasting a fat man who is stuck in the mouth of a cave (a party of potholers have imprudently allowed the fat man to lead them as they make their way out of the cave, and he gets stuck, trapping the others behind him. The right thing to do, clearly, is to sit down and wait until the fat man grows thin; but philosophers have arranged that flood waters should be

⁵² The example is of course H. J. McCloskey's ("An Examination of Restricted Utilitarianism" and a "Non-utilitarian Approach to Punishment").

While executing persons lies outside the prerogative of sheriffs (it lies within that of judges alone), McCloskey appears to believe otherwise. Also, being Australian, McCloskey's example is rather distasteful.

rising within the cave, and that the party have with them a stick of dynamite with which they can blast the fat man out of the mouth of the cave).⁵³

Returning to the question before us, recall first that moral dilemmas arise because the agent's moral beliefs imply that she ought to do both *A* and *B* where the circumstances are such that doing *A* necessitates refraining from *B*, and vice versa. For a state of affairs to constitute a moral dilemma then, it must consist in a (i) *pair* of (ii) *conflicting* actions or courses of action, both of which are (iii) *called for*.

Next note that the sheriff is faced with a choice between—on the one hand—preventing hundreds of deaths by killing an innocent, and—on the other hand—letting hundreds of people die by refraining from killing an innocent. The former is an act (a bodily movement or a set of bodily movements) that results in—or amounts to—both preventing hundreds of deaths and killing an innocent. The latter is refraining from that very act. Since refraining from an act is itself an act (this assumption is crucial for Chapter 2), the sheriff is faced with a choice between *two* acts: namely, an act and its negation. The scenario under consideration therefore meets condition (i). It also meets condition (iii), because both acts are—in one respect—called for (in another respect, both are “called against”). That the scenario meets condition (ii) has already been noted: committing and refraining from one and the same act is impossible. Thus, the scenario under consideration and its like certainly constitute moral dilemmas.

We should accordingly develop for cases of this sort an evaluation procedure analogous to the evaluation procedure for ‘or.’ You will recall that the evaluation procedure for ‘or’ is made up of a series of ‘if-then’ clauses such as the following.

⁵³ All three examples and more appear in Philippa Foot's “The Problem of Abortion and the Doctrine of Double Effect.”

If $v[A] + v[\neg B] > v[B] + v[\neg A]$, then

- (i) $v[A \vee B] = v[A]$
- (ii) $v[B \vee A] = v[B] + v[\neg A]$

These clauses are applicable to dilemmas containing right actions (both A and B enjoy a positive value). To see this, consider the clause before us. Line (i) reflects the notion that agents shouldn't be punished for maximizing rightness, for it requires that $v[A \vee B]$ is $v[A]$ and not $v[A] + v[\neg B]$. This is appropriate precisely because $v[\neg B]$ is negative: that is, because $v[B]$ is positive. Line (ii) reflects the notion that, even if failing to maximize rightness, agents should be credited for the rightness they do produce. Again, it's precisely because $v[B]$ is positive that line (ii) conveys this message.

However, the cases under consideration invariably contain a right and a wrong act. The reason is that they present agents with a choice between an act and its negation: that is, between A and $\neg A$. Thus, if A 's value is positive, $\neg A$'s value is negative, and vice versa. An evaluation procedure suitable for *such* dilemmas would presumably be the following.

- (i) $v[A \vee \neg A] = v[A]$
- (ii) $v[\neg A \vee A] = v[\neg A]$
- (iii) $v[\vee \neg A, A] = v[\neg A]$

Suppose that A 's value is positive. Then $\neg A$ value is negative: which means that opting for A amounts to maximizing rightness. Thus, the value of opting for it should be as right as A itself. This intuition is expressed in line (i). Now—as noted—if A 's value is positive, $\neg A$'s value is negative. Since opting for $\neg A$ excludes A , it involves production of no rightness whatsoever. By disregarding A 's value, line (ii) reflects this notion. Finally, since refraining from both horns amounts in effect to refraining from A , its value

equals the value of $\neg A$. Line (iii) conveys this information. Since however this information is superfluous, I shall henceforth disregard line (iii).

Note that if A 's value is *negative* (and $\neg A$'s value positive), lines (i)-(ii) remain the same. In particular, opting for $\neg A$ amounts to maximizing rightness, and should therefore be as right as $\neg A$ itself. Hence, $v[\neg A \vee A] = v[\neg A]$, as stated in line (ii). Opting for A on the other hand generates no rightness whatsoever. This means that opting for it shouldn't enjoy the positive effect of $v[\neg A]$. Hence, $v[A \vee \neg A] = v[A]$, as stated in line (i). Lines (i)-(ii) remain the same then regardless of whether $v[A] > v[\neg A]$ (i.e., regardless of whether A 's value is positive or negative). For this reason, the truth-table for $v[A \vee \neg A]$ and $v[\neg A \vee A]$ need not take a bifurcated 'if-then' format such as the following.

If $v[A] > 0$, then...
If $v[A] < 0$, then...

Also note that lines (i)-(ii) cohere perfectly with the truth-table I proposed in Section 10. First assume that A is preventing hundreds of deaths by killing an innocent, and $\neg A$ is allowing hundreds of people die by refraining from killing an innocent. Then,

$$(i) \quad v[A \vee \neg A] = v[A]$$

and

$$(ii) \quad v[\neg A \vee A] = v[\neg A]$$

Next let preventing hundreds of deaths and killing an innocent be $d_1\{A\}$ and $d_2\{A\}$, respectively. And let allowing hundreds of people die and refraining from killing an innocent be $d_3\{\neg A\}$ and $d_4\{\neg A\}$, respectively. Then,

$$(iii) \quad v[d_1\{A\} \text{ by } d_2\{A\}] = v[(d_1 + d_2)\{A\}]$$

and

$$(iv) \quad v[d_3\{\neg A\} \text{ by } d_4\{\neg A\}] = v[(d_3 + d_4)\{\neg A\}]$$

Finally note that as a matter of fact, A is given under $(d_1 + d_2)$ and $\neg A$ under $(d_3 + d_4)$. Thus.

$$(v) \quad v[(d_1 + d_2)\{A\}] = v[A]$$

and

$$(vi) \quad v[(d_3 + d_4)\{\neg A\}] = v[\neg A]$$

From (i), (iii), and (v) it follows that

$$(vii) \quad v[d_1\{A\} \text{ by } d_2\{A\}] = v[A \vee \neg A]$$

and from (ii), (iv), and (vi) it follows that

$$(viii) \quad v[d_1\{\neg A\} \text{ by } d_2\{\neg A\}] = v[\neg A \vee A]$$

Isn't that nice.

13. Conclusion

'By' is a major moral connective. It signifies identity, thus obtaining only between actions and themselves (usually under different descriptions).

The truth-table for 'by' is:

$$v[d_1\{A\} \text{ by } d_2\{A\}] = v[(d_1 + d_2)\{A\}]$$

meaning that the value of $\{A\}$ -under-a-description- d_1 by $\{A\}$ -under-a-description- d_2 equals the value of $\{A\}$ -under-both- d_1 -and- d_2 .

If $(d_1\{A\} \text{ by } d_2\{A\})$ by $d_3\{A\}$ and its like are wffs, the truth-table for 'by' is:

$$v[(d_1\{A\} \text{ by } d_2\{A\}) \text{ by } \dots d_n\{A\}] = v[(d_1 + d_2 + \dots + d_n)\{A\}]$$

In case A has both good and bad consequences,

$$v[d_1\{A\} \text{ by } d_2\{A\}] = v[A \vee \neg A]$$

$$v[d_1\{\neg A\} \text{ by } d_2\{\neg A\}] = v[\neg A \vee A]$$

Chapter 6: Concluding Remarks

1. Summary

In Chapters 2-5 I constructed a many-valued morality by identifying four major moral connectives and characterizing their syntax and semantics. A brief review of these characterizations may prove helpful.

'Not'

If A is a wff, then $\neg A$ too is a wff. But $\neg\neg A$ is *not* a wff. Nor is $\neg\neg\neg A$ or any string of symbols containing more than a single occurrence of \neg at a time.

Negation of action consists in refraining from it (recall that wffs in a many-valued morality are simple action descriptions). The valuation function for 'not' is:

$$\text{If } -.5 \leq v[A] \leq .5, v[\neg A] = -1.7 v[A]$$

$$\text{If } v[A] > .5, v[\neg A] = -1.7(1-v[A])$$

$$\text{If } v[A] < -.5, v[\neg A] = -1.7(1+v[A])$$

'Or'

If A and B are wffs, then $A \vee B$ too is a wff. However, $(A \vee B) \vee C$ or $[(A \vee B) \vee C] \vee D$ are *not* wffs. Rather, $A \vee (B, C)$, $A \vee (B, C, D)$, and so on are wffs.

$A \vee B$ reads 'choosing among a pair of incompatible acts, and attempting to act upon the choice.' Where A and B are both required, the evaluation procedure for \vee is:

If $v[A] + v[\neg B] > v[B] + v[\neg A]$, then

$$(i) \quad v[A \vee B] = v[A]$$

$$(ii) \quad v[B \vee A] = v[B] + v[\neg A]$$

If $v[A] + v[\neg B] < v[B] + v[\neg A]$, then

$$(i) \quad v[A \vee B] = v[A] + v[\neg B]$$

$$(ii) \quad v[B \vee A] = v[B]$$

If $v[A] + v[\neg B] = v[B] + v[\neg A]$, then

$$(i) \quad v[A \vee B] = v[A]$$

$$(ii) \quad v[B \vee A] = v[A]$$

If $v[\neg A] > v[\neg B]$, then

$$(i) \quad v[\mathbf{V} A, B] = v[\neg A]$$

If $v[\neg B] > v[\neg A]$, then

$$(i) \quad v[\mathbf{V} A, B] = v[\neg B]$$

The evaluation procedure for n -lemmas states that opting for the required—i.e., value maximizing—act enjoys that act's degree of rightness. Opting for the second (third, fourth, and so on) best option is judged with an eye to the required option alone.

For instance, if $v[A] = .5$, $v[B] = .6$, and $v[C] = .7$, then

$$(i) \quad v[A \mathbf{V} (B, C)] = v[A]$$

$$(ii) \quad v[B \mathbf{V} (A, C)] = v[B] + v[\neg A]$$

$$(iii) \quad v[C \mathbf{V} (B, A)] = v[C] + v[\neg A]$$

'And'

If A and B are wffs, then $(A \wedge B)$ too is a wff. However, as a matter of convention, I drop the outermost pair of brackets from a formula. Thus, $A \wedge B$ is a wff, as is $(A \wedge B) \wedge C$, $[(A \wedge B) \wedge C] \wedge D$, $(A \wedge B) \wedge (C \wedge D)$, and so on.

A conjunctive act consists in performing at least two distinct moral actions either concurrently or at different times. It amounts to revealing or exercising one's moral character.

Revealing or exercising one's moral character—a conjunctive action—is assigned the value of a person's moral character (which is her overall or global moral ability): that is, the weighted average of the values of the conjuncts. Thus, the evaluation procedure for conjunction is:

$$v[(A_1 \wedge A_2) \wedge \dots \wedge A_n] = n_i \bar{x}_i / (n_i + n_j) + n_j \bar{x}_j / (n_i + n_j)$$

'By'

If $d_1\{A\}$ and $d_2\{A\}$ are wffs, then $d_1\{A\}$ by $d_2\{A\}$ is a wff. I do not have a clear intuition concerning whether $(d_1\{A\}$ by $d_2\{A\})$ by $d_3\{A\}$ and its like are also wffs. In some moods I'm inclined to think that they aren't, for it appears to me that descriptions such as 'killing by shooting by pulling the trigger' or 'killing by shooting by pulling the trigger by moving one's finger' are hard or even impossible to parse. On other occasions however such descriptions seem to me almost crystal clear: which means that the corresponding symbolizations *are* wffs.

$d_1\{A\}$ by $d_2\{A\}$ reads: $\{A\}$ -under-a-description- d_1 by $\{A\}$ -under-a-description- d_2 .

The truth-table for 'by' is:

$$v[d_1\{A\} \text{ by } d_2\{A\}] = v[(d_1 + d_2)\{A\}]$$

meaning that the value of $\{A\}$ -under-a-description- d_1 by $\{A\}$ -under-a-description- d_2 equals the value of $\{A\}$ -under-both- d_1 -and- d_2 .

If $(d_1\{A\}$ by $d_2\{A\})$ by $d_3\{A\}$ and its like are also wffs, the truth-table for 'by' is:

$$v[(d_1\{A\} \text{ by } d_2\{A\}) \text{ by } \dots d_n\{A\}] = v[(d_1 + d_2 + \dots + d_n)\{A\}]$$

In case A has both good and bad consequences.

$$v[d_1\{A\} \text{ by } d_2\{A\}] = v[A \vee \neg A]$$

$$v[d_1\{\neg A\} \text{ by } d_2\{\neg A\}] = v[\neg A \vee A]$$

2. Moral and Logical Connectives

Each moral connective resembles to some degree a logical connective. Negation of actions echoes negation of sentences or propositions, for both signify an absence (the former of an action, the latter of a fact). Disjunction of actions mirrors exclusive disjunction of sentences or propositions, for both read: either A or B but not both. (Recall however that in propositional logic, disjunction is interpreted inclusively; it reads: either

A and/or *B*.) Conjunction of actions is akin to conjunction of sentences or propositions, as both signify a union of things.

These resemblances are rather obvious. Less obvious is the similarity between 'by' and biconditionals, a similarity due to which we may refer to 'by' as byconditional. Yet it too exists. For note that 'The state orphaned the boys by electrocuting their parents'¹ means that electrocuting the boys' parents *entails* or *implies* orphaning the boys. This observation should come as no surprise, given that 'by' signifies identity and that every wff entails itself. In other words, since electrocuting the boys' parents ($d_1\{A\}$) *is* orphaning the boys ($d_2\{A\}$), electrocuting the boys' parents implies orphaning the boys ($d_1\{A\} \rightarrow d_2\{A\}$).

Of course, since electrocuting the boys' parents *is* orphaning the boys, orphaning the boys also implies electrocuting their parents ($d_2\{A\} \rightarrow d_1\{A\}$). (This entailment is clearly peculiar: after all, the boys need not be orphaned by the state's electrocuting their parents: they may be orphaned in any of a number of means and methods. Attempts (to my mind successful) to explain away this and related peculiarities appear in Chapter 5, Sections 6-8.) Thus, $d_1\{A\} \leftrightarrow d_2\{A\}$, which means that 'by' resembles a biconditional.

3. (Semantic) Validity

In light of these resemblances between moral and logical connectives, it would be natural to ask whether many-valued morality resembles propositional logic also as far as validity is concerned.

Validity is a mark of correct reasoning. In valid lines of reasoning (arguments), the conclusion *follows* from the premises. Formally this means that an arguments is valid

¹ The example is Jonathan Bennett's (*Events and their Names*, 188).

if there exists no assignment of truth values such that the value of the conclusion falls below the lowest-valued premise. (In two-valued propositional calculus, this definition boils down to the following. The conclusion of a valid argument is true whenever its premises are.)

Many-valued morality doesn't lend itself to an analogous notion of validity. To see this, suppose that an argument of many-valued morality is valid if there exists no assignment of moral values such that the value of the conclusion falls below the lowest-valued premise. Then $A \wedge B$ doesn't entail A (the reason of course is that there are assignments of moral values such that the value of the conclusion falls below the lowest-valued premise. For instance, $v[A] = .5$, $v[B] = .7$. For then, $v[A \wedge B] = \frac{1}{2} (.5 + .7) = .6$; which means that $v[A]$ falls below $v[A \wedge B]$). Intuitively however, $A \wedge B$ does entail A : after all, $A \wedge B$ is a union of A and B . We must therefore reject the view that an argument of many-valued morality is valid if there exists no assignment of moral values such that the value of the conclusion falls below the lowest-valued premise.

In point of fact, we shouldn't expect many-valued morality to resemble logic as far as validity is concerned. I said that in valid arguments, the conclusion *follows* from the premises. This means that if a rational person endorses (stands in the belief relation to) the premises of a valid argument, she must also endorse its conclusion. Now if many-valued morality resembles logic as far as validity is concerned, the same applies to premises and conclusions of its arguments. But it doesn't. The reason is that while propositions—the subject matter of logic—are objects of belief, actions—the subject matter of many-valued morality—aren't. (One doesn't *believe that* shooting an elephant.)

In other words, the concept of validity is inextricably tied to the notion of reasoning (as is logic as a whole; for logic is nothing but the study of correct reasoning and the rules which govern it). Since actions play no role in reasoning (they aren't objects of belief), 'validity' is inapplicable to arguments comprised of them.

Having said that, note that due to equivalence relations between them,² there do exist six consequence-like relations between wffs of many-valued morality. Specifically, in case A has both good and bad consequences,

$$v[A \vee \neg A] = v[d_1\{A\} \text{ by } d_2\{A\}]$$

and

$$v[\neg A \vee A] = v[d_1\{\neg A\} \text{ by } d_2\{\neg A\}]$$

Therefore, $d_1\{A\}$ by $d_2\{A\}$ entails—and is entailed by— $A \vee \neg A$. Ditto for $d_1\{\neg A\}$ by $d_2\{\neg A\}$ and $\neg A \vee A$. (Four consequence-like relations so far.)

In addition, $A \vee B$ entails—and is entailed by— $A \wedge \neg B$ whenever one is willing to compromise certain moral intuitions. (All in all, six consequence-like relations.)

Many-valued morality then consists primarily in operations on objects. Therefore, despite the similarity between moral and logical connectives (operations), it is akin not so much to propositional logic, but to arithmetic instead.

4. Note on the Practical Syllogism

I said that actions play no role in reasoning. Yet in his *Nicomachean Ethics* Aristotle argues to the contrary. He suggests that practical syllogisms (of which ethical syllogisms are but a special case) have actions as their consequences. For example, if

(1) Everything sweet ought to be tasted

² Recall that A is equivalent to $B \dashv A \Leftrightarrow B$ —iff $v[A] = v[B]$

and

(2) This is sweet

then an agent who is capable of acting must not simply affirm

(3) This ought to be tasted

but rather actually act accordingly.³ (Aristotle doesn't provide examples of *ethical* practical syllogisms.)

I of course cannot here analyze and assess Aristotle's suggestion (for one thing, it must be analyzed and judged in the context of his overall ethical theory).⁴ Assuming however that it's defensible—assuming that is that (1) and (2) entail not (3) (a proposition), but rather an action—this suggestion sheds little if any light on the issue I discussed in the previous Section. This is because the practical syllogism has *propositions* for premises, whereas arguments of many-valued morality are comprised exclusively of actions.

³ Book 7, Chapter 3 (1147a).

⁴ For an in-depth treatment of the practical syllogism, see Chapter 1 of John Cooper's *Reason and Human Good in Aristotle*.

5. Calculus of Actions

Many-valued morality is a calculus of *moral* actions. Is there room for a continuum-valued calculus of actions (that is, actions in general, moral and non-moral alike)?

Two lines of thought suggest that we should answer this question in the affirmative. The first appeals to the concept of rationality. Assuming that an act is rational to the degree it effectively serves its purpose, moral and non-moral actions alike (but not *all* acts; explanation is forthcoming) enjoy a certain degree of rationality. Some actions are completely rational as they best serve their purposes, others enjoy lower degrees of rationality because they serve their purposes less effectively, and still others are completely *irrational* since they don't at all serve their purposes. (Should the rationality degree of actions be judged with an eye to only their subjective consequences? Or rather to only their objective consequences? Or perhaps to both kinds of consequences?)

Both moral and non-moral acts then may be assigned a value that represents a scalar property (namely, rationality). There's therefore a point in determining the rationality degree (value) of refraining from an act whose degree of rationality is such-and-such. Or the rationality degree (value) of choosing among incompatible acts, each of which enjoys a certain degree of rationality. In short, there is room for a continuum-valued calculus of actions in general.

As to why some acts don't enjoy a degree of rationality, I said that acts are rational in proportion to the degree they effectively serve their purposes. Arguably however, scratching one's head or biting one's lower lip while solving a problem are

purposeless actions; which means that these acts and their like are *non*-rational: they lie outside the rational realm. Consequently they also lie outside the scope of a calculus that relies on the concept of rationality. The first line of thought then suggests that there is room for a many-valued calculus of acts in general (moral and non-moral alike), not for a many-valued calculus of *all* acts.

Not so with respect to the second line of thought. A second route to a continuum-valued calculus of actions (in general) relies on consequentialism. Consequentialists hold that the moral value of acts depends (solely) on their consequences. Since all acts have consequences, all are right or wrong. Therefore, all acts lie within the moral realm. Hence, a many-valued calculus of moral acts willy nilly constitutes a many-valued calculus for all acts. In sum, the second line of thought involves the notion that since all acts lie within morality, a many-valued calculus of *moral* acts willy nilly constitutes a many-valued calculus for *all* acts.

6. Summary

Many-valued morality consists primarily in operations on objects. Therefore, despite the similarity between moral and logical connectives (operations), it is akin not so much to propositional logic, but to arithmetic instead.

Two lines of thought suggest that there is room for a continuum-valued calculus of actions (that is, actions in general, moral and non-moral alike). One such line revolves around the concept of rationality, the other around consequentialism.

Part II: Applications

In Part I of this essay I introduced one form of many-valued morality and explored its ramifications. Implications of just *any* many-valued morality are the subject matter of Part II.

In Chapter 7 I advance the claim that by contrast to both two- and three-valued moralities, a many-valued framework allows one to always take into account more than a single standard when evaluating actions.

Chapter 8 offers a sympathetic reading of Kant's 'lack of moral worth (content).' I maintain in it that from a many-valued perspective, the notion in question (which many deem problematic) is but a third moral value (in addition to 'right' and 'wrong'), and reflects an attempt to capture moral gradation.

Finally, in Chapter 9, I subject to many-valued analysis an aspect of the dispute between utilitarians and deontologists. I suggest in particular that whether killing a few to save many is unjust is a matter of degree: whether utilitarianism conflicts with justice over this issue is *not*.

Chapter 7: Pluralism or Multi-standard Morality

1. Introduction

We tend to measure actions against more than a single moral standard. In this Chapter I argue that while this tendency cannot be captured by existing—that is, two- and three-valued—moral systems, it *can* be accommodated by a many-valued morality.

2. The Tale of Abner and Gideon

Abner and Gideon, both of whom are employees of the same dot-com company (luftgeschäften.com, in case you care to know), share a passion for philosophy (each had a short career as a graduate student of philosophy before he grew smarter and turned to luftgeschäften). It is unsurprising therefore that they have forged a close friendship and spend much time together, not only next to the water fountain at the office, but also during after-work hours.

The other day, they entered a store with the intention of purchasing a pair of winter boots for Gideon. Not long after, Gideon witnessed Abner slipping a white, cheap cotton T shirt into his backpack: he later walked out of the store without paying for it, thereby removing all doubts about the nature of his act from Gideon's mind. Using both Kantian and consequentialist considerations, Gideon tried to persuade Abner to return the stolen item to its lawful owners at once (Gideon was a sensitive student of ethical theory, even though he specialized in philosophy of language; such things happen). To no avail.

Gideon found himself between a rock and a hard place. Since he values property laws and the notion of justice that they embody, he felt an obligation to disclose to the store owner that she had just been robbed of an item, cheap though it may be. Yet Gideon also values friendship, and he knew full well that turning Abner in to the store

owner (or even to the police; *that* thought, he later admitted to my face, has also crossed his mind) would terminate their friendship (which he kept valuing despite the petty theft: nobody's perfect, he recalled thinking to himself: I too engage in unbecoming behaviors—I'm a heavy smoker for instance—with which Abner puts up: why can't I come to terms with *his* shortcomings?).

I told you the story about Abner and Gideon (the end? Gideon never turned his friend in; but their friendship nonetheless dissolved quite rapidly after)—I told you that story in order to illustrate that one and the same act may promote friendship while undermining justice and vice versa (it's a matter of course that other values could have served my purposes equally well). If neither value translates into the other, and if—in addition to that—they don't both translate into a third value such as pleasure—in short, if friendship and justice are incommensurable values, they give rise to two distinct moral standards. According to the first, agents should promote friendship; according to the second they should promote justice.

A couple of comments before I move on. First of all, I don't wish to be read as suggesting that on each and every occasion, the only appropriate response to all moral values is promotion. As Christine Swanton argues, an appropriate response to friendship for example may be expression, honoring, or both.¹ However, for simplicity's sake, I shall focus on promotion.

The second remark concerns incommensurability of values (or value pluralism). The argument I am about to present in this Chapter certainly depends on there being more than a single moral standard against which actions are judged; but there being more than

¹ "Profiles of the Virtues."

a single standard does *not* depend on value pluralism.

Take the fact that we tend to evaluate actions on the one hand according to their objective consequences—namely, the effects which *in fact* ensue from them—and on the other hand the consequences foreseen by agents (subjective consequences). Objective and subjective consequences can, and often do, diverge considerably. Sara might intend to help her mother mow the lawn but end up interfering with the chore; Capone might have intended to terminate an innocent man but terminate Hitler instead (after all, Capone and Hitler *were* contemporaries).² This divergence indicates that objective consequences are distinct from subjective ones; thus, each gives rise to a distinct moral standard. Now divergence of objective and subjective consequences is possible even if we reject value pluralism and embrace in its place value monism such as hedonism (the theory that only pleasure is intrinsically valuable): agents may intend to generate pleasure but instead inflict pain; think of some instances of a massage. Therefore, multiplicity of moral standards is independent of whether some values are indeed incommensurable.

3. Two- and Three-valued Moralities

Two-valued moral systems are incapable of capturing our tendency to assess actions according to more than a single standard. Recall that a two-valued morality allows one to assign to actions just one of only two moral values, ‘right’ and ‘wrong.’ Next consider an act which is right according to one standard, but wrong according to

² Even though intentions constitute a subgroup of foreseen consequences, I use ‘intends’ and ‘foresees’ interchangeably. The reason is that my discussion concerns only instances where intentions *do* overlap with foreseen consequences.

another. Sara's act is a good example, for it's right when judged by its subjective consequences, wrong when judged by its objective outcomes. Gideon's act is also a good example, for it's right when assessed by its contribution to friendship, wrong when assessed by its contribution to justice. For brevity's sake, I shall confine my discussion to Sara's act. Assigning 'right' to Sara's act would be tantamount to evaluating it according to its subjective consequences alone. On the other hand, assigning her action the value 'wrong' is tantamount to judging it only in relation to its objective consequences. In sum, a two-valued framework allows one to say only that Sara's act is right (wrong), thereby imposing on one a choice in a single moral standard: namely, subjective (objective) outcomes. For this reason, a two-valued morality *completely* fails to represent the plurality of moral standards.

A three-valued framework is capable of capturing this plurality *to some extent*. A three-valued morality, you will recall, allows one to assign to actions any of three moral values, 'right,' 'wrong,' and 'neutral.' Now suppose that an action is right when judged by its subjective consequences, wrong when judged by its objective outcomes, and—in addition to that—right to the same degree that it is wrong. Sara's act would again be a good example. Another example would be really and truly intending and attempting to save a life but killing the intended beneficiary in the process. A three-valued moral framework has the resources for expressing what we wish to say, as it allows saying that these acts are neutral. But if an action is *very* right when judged by its subjective consequences and only *mildly* wrong when judged by its objective consequences, it would be inappropriate to say that it's neutral. The same applies of course to any action

which is right (wrong) according to one standard, wrong (right) according to another, and whose rightness degree differs from its degree of wrongness.

4. Many-valued Morality (1)

By contrast to both two- and three-valued systems, a many-valued morality can *fully* capture the plurality of moral standards. Suppose that an action is *very* right when judged by its subjective outcomes and only *somewhat* wrong when judged by its objective consequences. A many-valued system puts one in a position to articulate this by stating (for example) that the action—though clearly of a certain degree of rightness—could have been even *more right* had its objective consequences been better.

(I have heard qualms concerning this example. If the action had better consequences, it's been argued, it wouldn't be the same action. Perhaps. Ultimately my argument doesn't hang on this or similar examples.)

5. Many-valued Morality (2)

I should now like to establish the foregoing claims more formally. To achieve this end, certain habits ought to be unlearned. In the system I proposed in Part I, an action's moral value is ultimately determined according to its subjective consequences alone. Major moral intuitions motivate this assumption, but I ask that it now be set aside. Instead assume that actions are measured against two moral standards: their subjective and objective consequences.

(My request should generate no discomfort. As I have been suggesting throughout this work, a many-valued morality can take various shapes and forms. Systems may differ from each other in many respects, including the interval they employ.

the valuation procedures for the various connectives, or indeed the connectives themselves. They may also differ in the moral intuitions that they embody.)

You will recall that a two-valued framework allows one to assign to actions just one of two moral values, 'right' and 'wrong.' Replace 'right' and 'wrong' with the numerical values 1 and -1, respectively. Now consider Sara's act. Judged by its subjective consequences, its value is 1; judged by its objective consequences, its value is -1. What is the act's overall value? Presumably the average of these values. Since a two-valued framework allows one to assign actions either 1 or -1, the average—the act's overall value—must be either 1 or -1. If it's 1, it must be because $\frac{1}{2}(1 + 1) = 1$; which entails that, contrary to our hypothesis, Sara's act is right according to *both* its objective and subjective consequences. If on the other hand the act's overall value is -1, it must be because $\frac{1}{2}[(-1) + (-1)] = -1$. From *this* result it follows that, contrary to our hypothesis again, Sara's act is *wrong* according to both standards. One might suggest that an act's overall value need not be the *average* of its constituent values: instead it could be their arithmetical sum. This suggestion not only solves no problem, it exacerbates an existing one. $1 + (-1) = 0$. But 0 isn't among the wherewithal with which a two-valued framework furnishes one.

In sum, assigning to Sara's act 1 (-1) would be appropriate *only if* it were judged according to its subjective (objective) consequences alone. If it were evaluated according to both its subjective and objective outcomes—that is to say, according to more than a single moral standard—assigning to it the value 1 (-1) would misrepresent how we think. In other words, if the system captures the whole truth about Sara's act, the latter is assessed according to a single moral standard: if the act is assessed according to more

than a single standard, the system fails to capture the whole truth about it. Thus, a two-valued framework is capable of transmitting a faithful picture of only a single moral standard at a time. It is therefore *completely* incapable of capturing standard pluralism.

A three-valued framework is capable of capturing this pluralism *to a limited extent*. Recall first that a three-valued system allows one to assign to actions any of three values, 'right,' 'wrong,' and 'neutral.' Or 1, -1, and 0. Next consider Sara's act again. Judged by its subjective consequences, its value is 1; judged by its objective consequences, its value is -1. What is the act's overall value? 1, -1, or 0. There are no further options. As I argued earlier, if the act's overall value is 1, it's right according to both its subjective and objective consequences, contrary to our hypothesis. And if the act's overall value equals -1, it's *wrong* according to both its subjective and objective consequences, contrary to our hypothesis again. What if the act's overall value equals 0? Well, if that's the case, it must be because $\frac{1}{2} [1 + (-1)] = 0$; which implies that the act is right to the same degree that it is wrong. But if the act's rightness degree differs from its degree of wrongness, assigning to it 0 would misrepresent reality. A three-valued moral system then is capable of capturing the whole truth about Sara's act only if its rightness degree equals its degree of wrongness. Thus, such a framework is capable of accommodating the plurality of moral standards to only a limited extent.

(A three-valued system also allows for acts of value 1 (or -1) and 0 according to their subjective and objective consequences, respectively. To what extent is the system capable of accommodating *such* acts? Well, the overall value of such acts is 1, -1, or 0. If 1, it's because $\frac{1}{2} (1 + 1) = 1$; which entails that the act is right according to both its objective and subjective consequences, contrary to my assumption. If 0, it's because $\frac{1}{2}$

$(0 + 0) = 0$: which implies that the act is neutral according to both subjective and objective consequences, contrary to my assumption again. If -1 , this is because $\frac{1}{2} [-1 + (-1)] = -1$: which entails that the act is *wrong* according to both subjective and objective consequences, contrary to my assumption yet again. Put differently, $\frac{1}{2} (1 + 0) = \frac{1}{2}$. Since this value lies outside the scope of a three-valued morality, the latter is incapable of representing acts of the type under consideration. It follows that a three-valued moral framework is *completely* incapable of capturing such acts.)

By contrast to both two- and three-valued moral frameworks, a *many-* (or a continuum-) valued system is capable of *fully* capturing the plurality of moral standards. A many-valued framework allows one to assign to actions any real number between (in the present case) 1 and -1 . Now consider an act which is of value .8 according to its subjective consequences, and of value $-.3$ according to its objective consequences. In line with previous practice, the act's overall value equals to $\frac{1}{2} [.8 + (-.3)] = .25$. This result is by all means welcome. Since it's positive, it reflects the fact that the act's rightness degree exceeds its degree of wrongness. Since it's lower than .8, it reflects the fact that there is a sense in which the act is wrong.

6. Beyond Two Moral Standards

To make my case against two- and three-valued theories, I have employed a plurality of moral standards consisting of only two moral standards. Presumably however, there are as many moral standards as there are incommensurable moral values. It's worth noting that as the number of different moral standards one cares to consider increases, so does the inadequacy of two- and three-valued moralities. Throughout my

discussion I shall assume that an act's overall value always reflects as many standards as possible.

Take an act which is unusually right according to standard S_1 , highly right according to S_2 , and rather wrong according to both S_3 and S_4 . A two-valued morality has it that according to S_1, S_2, S_3, S_4 , the act's value is 1, 1, -1, and -1, respectively. The act's overall value could be either 1 or -1. If it's 1, this is because $\frac{1}{4}(1 + 1 + 1 + 1) = 1$. Such a result misrepresents two standards, namely S_3 and S_4 . If the act's overall value is -1, it is because $\frac{1}{4}[(-1) + (-1) + (-1) + (-1)] = -1$. This result too misrepresents two moral standards, namely S_1 and S_2 . When four moral standards are taken into account then, a two-valued morality may misrepresent two standards. (Of course, the system may misrepresent three and four standards as well. For example, it would misrepresent three standards when (i) according to S_1, S_2, S_3, S_4 , the act's value is 1, 1, 1, and -1, respectively, and (ii) the act's overall value is -1. However, as noted at the outset of this Section, I assume that an act's overall value always depicts a true picture of as many standards as possible. In the present case therefore, the act's overall value wouldn't be -1, but rather 1; in which case only a single standard—namely, S_4 —would be misrepresented.) But when only two moral standards are taken into consideration, a two-valued morality may misrepresent one of them at the most.

Similar claims apply to three-valued morality. Consider an act whose value is 1, 1, 1, -1, and -1 according to S_1, S_2, S_3, S_4, S_5 , respectively. Its overall value can be 1, -1, or 0. If 1, it's because $\frac{1}{5}(1 + 1 + 1 + 1 + 1) = 1$; a result that misrepresents two standards: S_4, S_5 . If -1, it is because $\frac{1}{5}[(-1) + (-1) + (-1) + (-1) + (-1)] = -1$; a result which misrepresents three standards: namely, S_1, S_2 , and S_3 . As for 0, the overall value

of the act under consideration can never equal 0; I take this to be obvious upon reflection. Thus, when five moral standards are taken into account, two of them may be misrepresented. (Recall that I assume that an act's overall value always paints a true picture of as many standards as possible.) However, if two moral standards are taken into account, a three-valued morality may misrepresent one of them at the most.

No such claims apply to a many- (that is, continuum-) valued morality. Its adequacy remains constant across numbers of distinct moral standards. The reason is probably obvious, but I shall nonetheless spell it out. The overall value of an act may equal any real number whatsoever between 1 and -1 . Since the interval of a continuum-valued morality *consists in* any real number between 1 and -1 , no overall value lies outside its scope. The number of values (i.e., the number of standards) employed to determine a given overall value doesn't change this one bit.

7. Objective Consequences Revisited

One might worry that Sara's act fails to serve my purposes. For unforeseen objective consequences—unfortunate and fortunate alike—do not affect the moral value of actions. Therefore, its bad outcomes don't make Sara's act *wrong* (though they do make it *bad*). Her act then is right and *bad*, not right and *wrong*. Contrary to my contention.

This objection is taken care of by an argument that draws on notions normally expressed in connection with the problem of determinism and freedom. Suppose that unforeseen objective consequences do affect the moral value of acts. Then it follows that an agent is morally responsible even when a significant aspect of what she does depends on factors beyond her control: if an agent has sound reasons to expect that his action

would have good (bad) consequences, but owing to some subsequent event which he could not foresee his action has bad (good) consequences, then his action is wrong (right). But this is surely incorrect. If factors which an agent could not have reasonably foreseen influence the consequences of his well- (ill-)intended action and make them bad (good), we are reluctant to say that the action is wrong (right) and that the person is blameworthy (praiseworthy). In general, clear absence of control, produced by involuntary movement, physical force, or ignorance of the circumstances, excuses what is done from moral judgment.

The main reason then for thinking that unforeseen objective consequences should not affect the moral value of acts is that such consequences are due to factors outside the agent's control, and persons should not be held morally responsible for what is due to factors beyond their control. The trouble is that when we start considering what is and isn't under our control, we come to realize that ultimately nothing or almost nothing is under our control. This point is usually made in connection with the paradox of determinism and freedom: people's intentions and character are not under their control: they have not chosen to be who they are or have the thoughts and desires that they have: all these were predetermined by genetic and environmental factors. Now if this is so—i.e., if we are not responsible for having the intentions that we have—then it makes no sense to blame or praise people not only for the actual (objective) consequences of their actions, but also for the intended (subjective) ones. This problem suggests that foreseen and actual consequences are on a par: if the former are morally (ir)relevant, so are the

latter.^{3 4}

8. Summary

When evaluating actions, we tend to employ more than a single moral standard. This tendency cannot be accommodated *at all* by a two-valued morality, and can be accommodated to only *a very limited degree* by a three-valued morality. By contrast to both two- and three-valued moralities, a many-valued morality is always capable of capturing the plurality of moral standards.

In addition to that, as the number of different moral standards one considers increases, so does the inadequacy of two- and three-valued moralities. By contrast, the adequacy of a many-valued system remains constant across numbers of distinct moral standards.

³ See Nagel, "Moral Luck," 580 and Murphy and Coleman, *Philosophy of Law*, 127.

⁴ Additional arguments for taking objective consequences into account appear in Peter Railton's "Alienation, Consequentialism and the Demands of Morality." Here is one. Achieving the best outcome sometimes requires *refraining* from intending to bring it about. A tennis player (to take Railton's simplest example) may come to realize that his obsession with winning is keeping him from playing his best: in order to play his best, he must *stop* thinking about winning. Now suppose that only subjective consequences (intentions) affect the moral value of acts. Then a person, who—in order to bring about the best outcome—refrained from intending to bring it, would be acting *wrongly*. Surely however, his conduct is right. And taking objective consequences into account paves the way for delivering just this judgment.

Chapter 8: Kant's 'Lack of Moral Worth'

1. Introduction

In the beginning of his *Groundwork of the Metaphysic of Morals*, Kant offers a series of four examples which imply that resentfully performing a dutiful action (saving the life of a person one loathes) has genuine moral worth, whereas lovingly or joyfully performing a similar action (saving the life of a loved one) altogether lacks moral worth. In Schiller's widely-cited words,

“Gladly I serve my friends, but alas I do it with pleasure.
Hence I am plagued with doubt that I am not a virtuous person.”
“Sure, your only resource is to try to despise them entirely,
And then with aversion to do what your duty enjoins you.”¹

Not surprisingly therefore, friends of Kant have found the notion of lack of moral worth quite disconcerting; and many of them have consequently embarked on attempts to accomplish a charitable reading of it.²

The present Chapter joins these efforts. I argue in it that far from troubling, 'lack of moral worth' is but a third moral value (in addition to 'right' and 'wrong') that reflects an attempt on Kant's part to capture moral gradation. Moreover, viewing Kant's examples from a many-valued (or scalar) perspective draws the sting out of Schiller's mockery. Thus, subjecting the notion of lack of moral worth to scalar analysis yields a sympathetic interpretation of Kant's views.

¹ *Über die Grundlage der Moral*, Section 6. Trans. by A. B. Bullock, quoted in H. J. Paton's *The Categorical Imperative*.

² Recent work includes Chapter 1 of Barbara Herman's *The Practice of Moral Judgment*, Keith Simmons's "Kant on Moral Worth" and Richard Henson's "What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action."

2. Preserving One's Life

Of the four examples offered by Kant to elucidate his thoughts regarding moral worth and lack thereof, the second and third are particularly instructive. The second example concerns preserving one's life.

...to preserve one's life is a duty, and besides this every one has also an immediate inclination to do so. But on account of this the often anxious precautions taken by the greater part of mankind for this purpose have no inner worth, and the maxim of their action is without moral content. They do protect their lives *in conformity with duty*, but not *from the motive of duty*. When on the contrary, disappointments and hopeless misery have quite taken away the taste for life; when a wretched man, strong in soul and more angered at his fate than faint-hearted or cast down, longs for death and still preserves his life without loving it—not from inclination or fear but from duty; then indeed his maxim has moral content.³

3. Helping Others

The third example concerns helping others.

To help others where one can is a duty, and besides this there are many spirits of so sympathetic a temper that, without any further motive of vanity or self-interest, they find an inner pleasure in spreading happiness around them and can take delight in the contentment of others as their own work. Yet I maintain that in such a case an action of this kind, however right and however amiable it may be, has still no genuinely moral worth. It stands on the same footing as other inclinations—for example, the inclination for honor, which if fortunate enough to hit on something beneficial and right and consequently honorable, deserves praise and encouragement, but not esteem; for its maxim lacks moral content, namely, the performance of such actions not from inclination, but *from duty*. Suppose then that the mind of this friend of man were overclouded by sorrows of his own which extinguished all sympathy with the fate of others, but that he still had power to help those in distress, though no longer stirred by the need of others

³ *Groundwork of the Metaphysic of Morals*, 65-6: 397-8. Italics his. (Again, page numbers refer to H. J. Paton's translation and the Prussian Academy of the Sciences edition, respectively.)

because sufficiently occupied with his own: and suppose that, when no longer moved by any inclination, he tears himself out of this deadly insensibility and does the action without any inclination for the sake of duty alone: then for the first time his action has its genuine moral worth.⁴

4. Three-valued Morality

Kant then distinguishes two classes of acts in accordance with duty (dutiful acts). One class consists in acts that have genuine moral worth. Their mark: performing them requires overcoming an inclination (desire) contrary to duty. Helping others or preserving one's life when one is not inclined to do so—when one is tempted to *refrain* from the act due to (say) all-pervasive desperation—has true moral worth. I shall henceforth refer to such acts as morally right.

A second remark concerning right acts is perhaps unnecessary, but—to be on the safe side—I shall make it nonetheless. Kant argues that since no supporting inclination impels their performance, morally right acts are motivated by duty alone. In Kant's view then, duty and inclination are distinct motivational forces which together exhaust the possibilities of motives for action.

The second class includes acts that altogether lack moral worth. They are characterized by the presence of an inclination in accordance with duty. Helping others when one is inclined to do so—perhaps because one takes pleasure in helping the needy—lacks moral worth. Donating a kidney to a loved one similarly lacks genuine worth.

Not only it would never occur to us to forbid acts belonging to the second category, they in fact deserve praise and encouragement.⁵ Clearly therefore, these acts

⁴ *Ibid.*, 66: 398. Italics his.

⁵ *Ibid.*, *ibid.*

are not wrong. Nor are they right, however: they *lack* moral worth. The acts in question then are neither right nor wrong. I shall accordingly designate them ‘neutral.’⁶

In a one-sentence remark preceding the four examples Kant identifies yet a third class of actions: namely, acts contrary to duty. He argues that since such acts are directly opposed to duty, the question whether they were motivated by duty alone doesn’t even arise. I take it that actions belonging to the third category are morally wrong.

We may legitimately hold then that Kant advances a three-valued morality. It allows one to assign actions either of three moral values—‘right,’ ‘neutral’ and ‘wrong’—depending on whether the act (i) accords with duty, and (ii) is motivated (also or only) by an inclination in accordance with duty. ‘Also or only’ because a dutiful act need not be motivated by exclusively duty or exclusively an inclination in accordance with duty: it may also be propelled by a configuration of both forces. Following Richard Henson,⁷ I shall call acts that are so impelled ‘overdetermined.’

The table summarizes Kant’s framework. (For simplicity of presentation, I assume that absence of inclination in accordance with duty amounts to presence of inclination contrary to duty. Though clearly defensible, this assumption is not indisputable. Nor is it necessary for my argument however.)

⁶ Further support for this move may be found in the later *Metaphysic of Morals*, where Kant maintains that deficiency in moral worth is tantamount to the value 0 (153: 6:390). (Page numbers refer to Mary Gregor’s translation and the Prussian Academy of the Sciences edition, respectively.)

⁷ “What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action.”

Dutiful	Yes	No
Inclination		
Yes	Neutral	
No	Right	Wrong
	a	b
	c	d

Fig. 8.1. Kant's Three-valued Morality

(Curiously enough, Kant's three-valued morality echoes the tripartite distinction—found in Judaism—between (i) failing to fulfill a *mitzvah*, (ii) fulfilling a *mitzvah* for its own sake, and (iii) fulfilling a *mitzvah* for the sake of some external benefit. While (i) is wrong, (ii) and (iii) are not; yet (ii) is superior to (iii).)

5. Or perhaps a Four-valued Morality?

The coordinate system employed by Kant for identifying the moral value of acts gives rise to a logical space consisting of not three—but rather *four*—types of acts. Does Kant's position entail that cell (b) must remain uninhabited? Could an act whose performance is motivated (also or only) by an inclination in accordance with duty be contrary to duty?

According to Kant, agents are motivated by duty if and only if they are motivated by reverence for (the formula of universal law of) the Categorical Imperative.⁸ You will recall that the Categorical Imperative states that if willing an action is consistent with willing that everyone under the same circumstances would do the same, the act is permissible. On the other hand if willing an act is *inconsistent* with willing that everyone

⁸ *Groundwork of the Metaphysic of Morals*, 68-70; 400-2.

in the same situation would do the same, the action is forbidden. I take it that an act is obligatory if willing its omission is inconsistent with willing that everyone under the same circumstances would omit the act.

Since respect for the Categorical Imperative presumably includes applying it—as a permissibility and obligatoriness test—to (principles of) actions and acting on the test's results, it guarantees that agents perform dutiful acts and refrain from forbidden ones. Whenever duty plays a motivational role then, dutiful conduct ensues. Hence, overdetermined acts—acts whose performance is motivated by *both* duty and inclination in accordance with it—could not possibly oppose duty.

On the other hand, acts whose performance is motivated *only* by an inclination in accordance with duty could well oppose duty. Acknowledging this possibility, Kant writes:

For if any action is to be morally good, it is not enough that it should *conform* to the moral law—it must also be done *for the sake of the moral law*: where this is not so, the conformity is only too contingent and precarious, since the non-moral ground at work will now and then produce actions which accord with the law, but very often actions which transgress it.⁹

Barbara Herman (in whose influential interpretation of Kant 'precarious' and its antonyms take center stage) provides an example:

Suppose I see someone struggling, late at night, with a heavy burden at the backdoor of the Museum of Fine Arts. Because of my sympathetic temper I feel the immediate inclination to help him out. We need not pursue the example to see its point: the class of actions that follows from the inclination to help others is not a subset of the class of right or dutiful actions.¹⁰

⁹ *Ibid.*, 57-8: 390.

¹⁰ *The Practice of Moral Judgment*, 4-5.

Does the fact that cell (b) is occupied, that in Kant's view there exists a fourth class of actions, suggest that his system constitutes not a three- but rather a *four*-valued morality? I think not. Since Kant carefully distinguished between cells (a) and (c), it stands to reason that he would have distinguished between cells (b) and (d) had he considered the move advantageous. Yet there appears to be no textual evidence to the effect that Kant distinguished between cells (b) and (d): for him, both cells are populated by acts that oppose or transgress duty and are therefore wrong. This means that according to Kant, acts may be assigned either of *three* moral values ('right,' 'neutral' and 'wrong.')

Which in turn means that he espouses a three-valued system.

Insisting on this conclusion is however quite unnecessary, as no major claim of mine hangs on it: none of my central contentions would be different had Kant advocated not a three- but rather a four-valued morality. I clarify this comment as I go along.

5. Why Three-valued Morality?

Having established *that* Kant offers a three-valued framework, let us now turn to the *why*: Why did Kant find it fit to introduce a third moral value in addition to 'right' and 'wrong'?

Begin by considering two complementary examples, both of which mimic Kant's own. Moishe is a well-to-do venture capitalist, whereas Avrum is a destitute janitor. They each donate \$200 to famine relief. While Moishe acts effortlessly (his donation requires no sacrifice on his part), Avrum is strongly inclined to refrain from the act (it requires forgoing a much needed home appliance). It seems fair to say therefore, that while Moishe's act is undoubtedly right, Avrum's is *unusually* so.

Moishe and Avrum both refrain from robbing a bank. Moishe refrains effortlessly (cash is anyway coming out of his ears), whereas Avrum refrains following a bitter battle against the temptation to substantially and speedily supplement his income (his son is about to enroll in nursing school). Again, it's fair to say that Moishe's refraining is right, though not to an extraordinary degree. On the other hand, Avrum's refraining merits—to employ Henson's terminology—a battle citation.

These examples embody the intuition that as the sacrifice or effort involved in performing a good action increases, so does the action's degree of rightness or moral worth. Ditto *Kant's* examples. Preserving one's life under circumstances of good health and relative affluence is certainly fine (there's nothing *wrong* about such conduct), yet doing so under more or less contrary circumstances—given severe disability for instance—is laudable. Helping others effortlessly is again just fine, whereas doing so when one's mind is “overclouded by sorrows of his own” is admirable.

Kant then perceives a moral difference between (i) dutiful acts whose performance requires no sacrifice on the agent's part and (ii) dutiful acts whose performance involves a battle against contrary inclinations. Since calling both types of acts ‘worthy’ would blur that difference, he reserves ‘worthy’ for the latter alone: to refer to the former, he introduces a third moral value: namely, lack of moral worth.

I am suggesting in other words that Kant introduces a third moral value in order to express the observation that worth wise, some dutiful acts are superior to others. Or indeed, to accommodate the view that moral worth admits of degrees.

Again, saving the life of a starving Somali at \$20 per month is certainly right, yet saving a life at risk of one's own life is extraordinarily so: which means that rightness

admits of degrees. As his examples show, Kant shares this observation. It is in order to accommodate it that he advances a three-valued morality.

6. Kant's Three-valued Morality is Too Weak

Three problems afflict Kant's three-valued moral theory. First of all, it is too weak in that it fails to reflect existing moral differences: it entails that some acts are equally worthy even when intuitively they differ. Secondly and complementarily, the theory is too strong in that it inflates moral differences: it implies that some acts *substantially* differ from others even when the difference is not all that dramatic. Finally, the theory implies that (say) saving the life of a loved one altogether lacks moral worth. These problems occupy the bulk of the remainder of this Chapter.

To appreciate the first problem, consider Golda and Hinda. Golda, a single mother and sole provider of two young children, is an assistant professor of Sociology at a community college. She can't help but notice that in order to maintain a reasonable graduation rate in the institution, her colleagues and supervisors engage in unbecoming practices. Winning a bitter battle against the inclination to secure the future of her offspring, she blows the whistle. Hinda, a twenty year old mother of six month old twins, is watching a movie on television. During a pivotal moment in the film, one of her babies bursts into tears. Winning a battle against the inclination to watch the movie, she approaches the baby and checks on her needs.

Having won a harder battle, Golda's conduct is more laudable than Hinda's. Yet Kant's theory entails that since both conducts involve overcoming an inclination contrary to duty, they are both right (right, period), and therefore of equal moral value. The theory thus conceals the difference between Golda's victory and Hinda's.

The difficulty is exacerbated as we turn our attention to forbidden acts (or acts contrary to duty). Impressed by the notion that moral worth is a function of sacrifice, Kant draws a distinction between dutiful acts that do and don't require sacrifice on the agent's part. Drawing an analogous distinction between forbidden acts is accordingly called for, since their values too are affected by sacrifice. As I argued in Chapter 2, Section 11, as the temptation to do a wrong increases, the act's degree of wrongness decreases. To repeat an example I gave there, if an agent grew up in an underprivileged environment, he would in all likelihood be very tempted to steal; and if he *were* to steal, his life circumstances are likely to be looked upon as mitigating circumstances as well. If on the other hand the agent grew up in an affluent surrounding, he is likely to be less tempted to steal; and if he were to steal, his life circumstances would be regarded as aggravating circumstances too. Since the temptation to do a wrong is presumably nothing but the sacrifice or effort required for refraining from an act, wrongness—like rightness—is a function of effort.

Kant's views then call for a distinction between forbidden acts that do and don't involve temptation. Yet as I argued in Section 4, Kant draws no distinctions whatsoever within the class of forbidden acts. His theory thus implies that *all* forbidden acts are equally wrong. Thus, while revealing at least *some* differences among dutiful acts, Kant's theory represents *no* difference whatsoever among forbidden acts.

The objection that Kant's three-valued theory is too weak can be cast in terms of information loss. In Kant's system, all right, neutral and wrong actions are indiscriminately categorized as right, neutral and wrong. We thus lose information about

the moral difference between bitter and easy battles against inclinations contrary to duty, between strong and weak temptations to transgress duty, and so on.

The objection can also be formulated in terms of scalar morality. Kant believes that moral worth is a function of the sacrifice or effort involved in performing an act. By recognizing only acts that do and don't involve sacrifice, he implies that sacrifice is an all-or-nothing (binary) property. Being a form of pain however, sacrifice is a *scalar* property. This means that a finitely-valued theory is incapable of fully capturing degrees of moral worth.

Note that being finitely-valued, a four-valued theory is equally vulnerable to the objection at hand.

7. And Too Strong

Kant's three-valued theory entails that helping others when one's mind is "overclouded by sorrows of his own" is right, whereas helping out of love or sympathy altogether lacks moral worth. The theory thus implies that these acts are *dramatically* different from each other. Intuitively however, the difference is not as dramatic: for intuitively, both instances of helping others are right, though right to different degrees.

One might complain that if both instances of helping others are right, there is no need to discard three-valued morality: merely replacing 'neutral' by 'slightly right' would do the trick. (Clearly, subscribing to a three-valued morality is perfectly consistent with holding that *all* dutiful acts possess moral worth.) In other words, one might argue that Kant advocates two independent theses. The first is that morality is three-valued. The second—that some dutiful acts altogether lack moral worth (or are neutral). Inflated moral differences are a consequence neither of the first thesis nor of the first and second

theses taken together. Rather, it's a consequence of the second thesis alone. Amending the second thesis should therefore solve the problem at hand.

Not so. It is the first thesis—and it alone—which is responsible for the problem of inflated differences. To see this, first replace 'slightly right' for 'neutral.' Next consider Golda and Hinda again. Assigning their respective conducts 'right' and 'slightly right' seems appropriate. However, if Hinda's battle were half as hard as Golda's (imagine that she battles with an inclination not to watch a movie, but to meet a crucial deadline), then assigning her act 'slightly right' would greatly inflate the difference between her conduct and Golda's. (On the other hand, assigning both conducts 'right' would conceal the disparity between them.)

Again, suppose that *A*, *B* and *C* are dutiful acts that require battling with inclinations of strength 10, 5 and 0, respectively. Assigning *A* and *B* 'right' and *C* 'slightly right' would inflate the disparity between *B* and *C* (and conceal the difference between *A* and *B*). Assigning *A* 'right' and *B* and *C* 'slightly right' would inflate the difference between *A* and *B* (and conceal the difference between *B* and *C*).

Unlike the previous problem, the present one does *not* intensify when we attend to forbidden acts. The previous problem worsens for forbidden acts because Kant's theory accurately represents (doesn't conceal) at least *some* differences within the class dutiful acts but *no* differences among forbidden acts. Due to entirely denying moral worth to denizens of cell (a), the theory inflates *all* differences within the class of dutiful acts: *no* disparity receives faithful representation; since Kant's theory also inflates all differences within the class of forbidden acts, they and dutiful acts are equally affected by the problem at hand.

It's a matter of course that the problem of inflated differences is totally independent of whether there exists a distinction within the class of forbidden acts. It would therefore afflict Kant's framework even if he introduced a fourth moral value to refer to the acts inhabiting cell (b).

8. Third Problem: Neutral Dutiful Acts (Schiller's Mockery)

I said that Kant's examples embody the intuition that as the sacrifice or effort involved in performing a good action increases, so does the action's degree of rightness or moral worth. This intuition has not been *uniformly* embraced by moral philosophers. (You will recall that Aristotle for example held that the virtuous individual accurately perceives what ought to be done in a given situation, and acts upon his perception rather effortlessly.¹¹) At the same time however, it's by no means perverse or eccentric. On the contrary, it's rather conventional.¹² There is therefore no reason why Kant should be beaten on the head for advocating the view that moral worth is a function of the sacrifice involved in performing a dutiful act.

There is however a reason for criticizing Kant for holding that (say) saving the life of a loved one altogether lacks moral worth. Entirely denying moral worth to dutiful acts whose performance is motivated (also or only) by an inclination in accordance with duty is indeed disturbing.¹³ It appears that such acts, though inferior to acts done from the motive of duty alone, are nonetheless right.

¹¹ Recall however that contemporary virtue ethicists are ambivalent with respect to this position of Aristotle's. See Chapter 2, Section 4.

¹² See for example Section 6 of Judith Jarvis Thomson's "A Defense of Abortion." Or Frances Kamm's "Supererogation and Obligation." Or Holly Smith's "Varieties of Moral Worth and Moral Credit."

¹³ Kant appears to manifest some awareness of this. While denying moral worth to them, he refers to denizens of cell (a) in unambiguously positive moral terms. See especially his remarks concerning helping others out of sympathy in Section 3.

As opposed to the previous problem, the present one *would* be solved by merely replacing ‘neutral’ by ‘slightly right.’ Recall that Kant espouses two independent theses. The first is that morality is three-valued. The second—that dutiful acts whose performance involves a supporting inclination entirely lack moral worth. Since the present problem is but a complaint against the second thesis, revising the latter alone should suffice for solving the problem.

Since this problem is unrelated to the number of moral values employed in a framework, it would afflict Kant’s theory even if he introduced a fourth moral value to refer to the denizens of cell (b).

9. Three- and Continuum-valued Moralities

Kant was very probably familiar with ancestors of three-valued formal frameworks, but certainly not familiar with continuum-valued logics. Specifically, the notion of a third truth-value appeared on the scene already in the Classical era in connection with the view—often attributed to Aristotle—that while statements about the past and present are unalterably true or false, statements about the future—such as ‘There will be a sea battle tomorrow’—are neither unalterably true nor unalterably false. Continuum-valued systems by contrast were introduced—by Łukasiewicz—only in the first half of the 20th century.¹⁴

I should now like to suggest that if continuum-valued frameworks had been available at the time, Kant would model his moral theory after *them*, not after (ancestors of) a three-valued system. I proceed by listing two (slightly overlapping) broad types of considerations to the effect that Kant’s views call for a continuum-valued morality.

¹⁴ See Williamson, *Vagueness*, 102, 114-20.

Considerations of the first type revolve around the fact that Kant construes evaluative concepts in terms of a scalar property (sacrifice). Considerations of the second type focus on the problems I presented in Sections 6-8. As one might expect, some of these considerations were already alluded to or anticipated in previous sections.

10. Scalar Rightness and Wrongness

Kant holds that dutiful acts are right if they require sacrifice on the agent's part, neutral if they don't. He thus implies first that an act's moral worth is a function of the sacrifice involved in its performance. Secondly he implies that sacrifice is an all-or-nothing property: acts either require it from agents or not. Being a form of pain however, sacrifice admits of infinitely many degrees (it constitutes a continuum). Which means that in order to accommodate Kant's views on moral worth and its determinants, a continuum-valued framework is called for.

Kant's position calls for infinitely many degrees of rightness, because he holds that an act's degree of rightness is a function of the effort involved in—or the temptation to refrain from—its performance. Since—as I explained in Section 6—an act's degree of *wrongness* too is affected by effort, Kant's views require infinitely many degrees of wrongness as well. Again then, to accommodate his implied stand on wrongness, Kant ought to adopt a continuum-valued framework.

In a word, since Kant's explicit or implied views on rightness and wrongness entail that they are scalar properties, his views require a continuum-valued morality.

11. Problems Afflicting Kant's Theory

Another reason why Kant's views call for a continuum-valued morality is that it would eliminate—in the first two cases directly and in the third one indirectly—the problems afflicting his three-valued morality.

First Problem: Concealed Differences Since in Kant's (explicit or implied) view rightness and wrongness are a function of a scalar property (sacrifice), they too are scalar. A finitely-valued framework is too weak for capturing scalar rightness and wrongness, as it entails that acts are morally equal even when they aren't. By contrast, a continuum-valued theory assigns acts their exact values, thereby expressing each and every disparity between acts. Adopting a continuum-valued morality would therefore solve the first problem afflicting Kant's three-valued theory.

Second Problem: Inflated Differences Kant implies that rightness and wrongness are scalar properties. A finitely-valued system imposes categorizations that inevitably inflate moral differences between acts. By contrast, a continuum-valued framework assigns each act its exact value, thereby accurately representing the differences between acts. A continuum-valued framework would thus solve the second problem afflicting Kant's three-valued morality.

Third Problem: Neutral Dutiful Acts According to Kant's theory, the acts inhabiting cell (a) altogether lack moral worth. Intuitively however, such acts—though inferior to denizens of cell (c)—are nonetheless right.

While each of the previous problems is a *consequence* of Kant's theory being finitely-valued, the present one isn't. Adopting a three-valued morality is consistent with holding that *all* dutiful acts are right. This means that espousing a continuum-valued

morality is by no means *essential* for solving the third problem. However, it would have probably helped Kant to avoid the problem in the first place.

Suppose that Kant embraced a continuum-valued morality—a framework with infinitely many degrees of rightness for dutiful acts and infinitely many wrongness degrees for forbidden acts—and at the same time reserved ‘lack of moral worth’ or 0 to a subset of dutiful acts. Then his position would be asymmetrical: for it would state that while all forbidden acts are wrong, not all dutiful acts are right. Kant would probably have noticed the asymmetry and consequently would have recognized the inappropriateness of entirely denying rightness to dutiful acts. In other words, if Kant adopted a continuum-valued morality, he would in all likelihood have noticed that since ‘lack of moral worth’ is neutral between dutiful and forbidden acts, it shouldn’t be assigned to dutiful acts.

12. Note on the Formula of Universal Law

As I argued in Chapter 1,¹⁵ Kant’s absolutism is incompatible with the notion that rightness and wrongness admit of degrees. You will recall that since in Kant’s view the deontic status of acts rests on logical contradiction, permissibility and obligatoriness are binary properties. As moral dilemmas effectively demonstrate however, obligatoriness is a scalar property. Moreover, both obligatoriness and rightness admit of degrees because goodness does. Which means that obligatoriness is a scalar property iff rightness is. Or that if obligatoriness is a binary property, so is rightness.

Kant’s absolutism then doesn’t sit well with his sensitivity to degrees of moral

¹⁵ Section 3.

worth. Note that in order to reach this conclusion, it's unnecessary to submit that Kant's views entail that rightness admits of *infinitely* many degrees. Suffice it to recognize that Kant acknowledges *two* degrees of rightness: namely, all and nothing. In particular, you will recall that if a property *P* is binary, the objects to which it applies are all equally *P*. Since according to Kant's absolutism obligatoriness—and consequently rightness—are binary properties, it entails that *all* dutiful acts are equally right. Therefore, Kant's absolutism conflicts with the notion that some dutiful acts are right while others are neutral. With the notion that is that there exist two degrees of rightness: all and nothing.

This argument serves to reply to two related worries. The first is that Kant's absolutism is compatible with two—but not with infinitely many—degrees of rightness. The second is that as a result, there is no ground for thinking that Kant's absolutism conflicts with his sensitivity to degrees of rightness.

One might attempt to restore consistency in Kant's views by affirming that obligatoriness doesn't admit of degrees while rightness does, and denying that goodness and rightness are correlated with obligatoriness (that is, conceding that obligatoriness and rightness are orthogonal or non-interacting dimensions). This attempt however is doomed to fail. The position that obligatoriness and rightness are orthogonal dimensions implies that there exist required (or permissible) but wrong and forbidden but right acts. While in and of itself plausible, this view is indefensible relative to Kant's opinions. In particular, where the deontic status of acts is legally defined, required (or permissible) but wrong and forbidden but right acts generate no conceptual dissonance. One need only consider Nazi Germany to appreciate this point. For Kant however, the deontic status of acts is determined by the (formula of universal law of the) Categorical Imperative. It

stands to reason that if the latter outputted forbidden but right or required (permissible) but wrong acts, it would never have made it into the *Groundwork*.

Kant's three-valued morality then conflicts with his sensitivity to degrees of moral worth. This conflict constitutes yet another reason—if another was needed—for rejecting Kant's absolutism.

13. Summary

Kant's introduction of 'lack of moral worth' was motivated by an attempt to capture the observation that rightness is scalar. It is the desire to accommodate the notion that some dutiful acts are superior to others that propelled his three-valued morality.

If continuum-valued formal frameworks had been available at the time, Kant would in all likelihood have modeled his morality after them, and not after (ancestors of) three-valued frameworks. This hypothesis gains credibility due to several considerations. In the first place, Kant understands rightness in terms of sacrifice; since sacrifice is scalar, rightness too is scalar; and scalar rightness calls for a continuum-valued morality. Secondly, Kant's views require infinitely many degrees of not only rightness but of wrongness as well; for an act's degree of wrongness is also affected by sacrifice. Finally, adopting a continuum-valued morality would correct two major defects in Kant's theory and would probably spare him a third shortcoming.

Seen from a scalar perspective then, Kant's theory is not at all troubling. Instead, it amounts to a significant step towards capturing moral gradation.

Chapter 9: Utilitarianism and Deontological Justice

1. Introductory Remarks

One of the two most influential objections that have been leveled against classical utilitarianism (henceforth, utilitarianism) over the years is that it is at odds with justice.¹ In what follows I examine this objection through many-valued lenses.

2. Setting the Stage

Recall the scenario I introduced in Chapter 5, Section 12. In a small, God forsaken American town a white woman has been raped and cruelly murdered. White residents of the town suspect that a black person committed the crime. Outraged and aching, they are about to take the law into their hands and lynch hundreds of black people. The local sheriff can prevent these killings by executing an innocent man as a scapegoat. Should he?

Utilitarianism requires an agent to always perform that act which, of all the alternatives open to her, would maximize satisfactions of (rational?) desires.^{2 3} It therefore entails that the sheriff should kill the innocent man. True enough, in actual cases of this sort the utilitarian would normally be in a position to bring her theory in line with common moral opinions. She will be able to point out for example that there is some possibility of the sheriff's dishonesty being found out, with consequent weakening of confidence and respect for the law in the community, the implications of which would be far worse than the painful deaths of hundreds of people. However, the case can be

¹ The other influential objection being of course that utilitarianism is excessively demanding.

² While this is a common reading of utilitarianism, recall that both Bentham and Mill provided also many-valued formulations of the doctrine (see Chapter 1, Section 4).

³ It is common to distinguish satisfaction of *rational* desires—*eudaimonia*—from satisfaction of desires *simpliciter* (the latter are usually designated 'happiness' or 'pleasure.'). Many believe that Bentham's utilitarianism employs the latter notion, whereas Mill's employs the former.

modified so that these objections don't apply: it could be imagined for instance that the sheriff could have first-rate empirical evidence that he will not be found out. That is to say, the scenario could always be strengthened to the point where we would have to admit that if utilitarianism is correct, the sheriff ought to execute the innocent man.

Authors with Kantian leanings believe that killing an innocent is unjust, come what may: they accordingly conclude that utilitarianism is at odds with justice.⁴ Philosophers of the utilitarian persuasion defend their theory, arguing that it in effect requires opting for the lesser of two evils.⁵ And members of a third group feel the force of both justice and utilitarianism: they hold that as the difference between the number of beneficiaries and the number of victims increases, so does the appeal of utilitarianism, while that of justice decreases. They hold in other words that as the difference in question increases, the degree of injustice involved in killing an innocent decreases, and—consequently—so does the tension between utilitarianism and justice.⁶ According to members of the third group then, the disharmony between utilitarianism and justice is a matter of degree.

Or is it?

To begin with note that if utilitarianism indeed requires agents to kill a few in order to save many, then such an act is not unjust *from a utilitarian point of view*. I mean: utilitarianism doesn't conflict with *its own* notion of justice. Arguing therefore that utilitarianism is incompatible with justice requires specifying what notion of justice

⁴ E.g., Elizabeth Anscombe, "Modern Moral Theories," 362-3 or H. J. McCloskey, "An Examination of Restricted Utilitarianism" and "A Non-utilitarian Approach to Punishment."

⁵ E.g., J. J. C. Smart in his "An Outline of a System of Utilitarian Ethics," 69-72 or Kai Nielsen, "Against Moral Conservatism," especially in the concluding Section.

⁶ E.g., Thomas Nagel, "War and Massacre," especially the concluding Section, and Philip Pettit, "Utilitarianism," 152-3.

is at work.

This notion is of course the deontological concept of justice, according to which producing the best results is sometimes forbidden. (This intuition is often captured in terms of constraints⁷: there are constraints on promoting the best outcome. One may not for example kill an innocent, non-threatening person for his organs to save five other lives.) I shall naturally focus on John Rawls's characterization of this concept.

3. Rawls 101

In Rawls's view, the principles of justice are the principles that free, rational, and self-interested persons would accept in an initial (or original) position of equality as defining the fundamental terms of their association.⁸ The original position is a hypothetical situation whose essential feature is that the persons in it are ignorant of their most socially significant characteristics: socio-economic status, natural abilities (physical and mental), race, gender, sexual orientation and identity, religion and the like. The principles of justice then are chosen behind a veil of ignorance. This ensures that no one is in a position to design principles that would favor his own condition.⁹

Rawls argues that persons in the original position would select two rather different principles. The first states that each person is to have an equal right to the most extensive liberty compatible with a similar liberty for all (the principle of equal liberty).

⁷ Ronald Dworkin refers to them as trumps (*Taking Rights Seriously*, xi), Thomas Nagel as deontological reasons ("Autonomy and Deontology," 143), and Robert Nozick as side-constraints (*Anarchy, State and Utopia*, 29).

⁸ *A theory of Justice*, 11.

⁹ *Ibid.*, 12.

The second says that social and economic inequalities are to be arranged so that they are both (a) to the benefit of the least advantaged (the difference principle), and (b) attached to positions and offices open to all (the principle of equal opportunity).¹⁰

While these principles *would* be chosen behind a veil of ignorance, the principle of utility would *not*. For it's highly unlikely that self-interested persons who view themselves as equals would agree to a principle which may require losses for some simply for the greater sum of advantages enjoyed by *others*.¹¹ (Not only *classical* utilitarianism would be rejected by persons in the initial position. Its sophisticated younger sibling—that is, *average* utilitarianism, which requires maximizing not the aggregate good but rather utility *per capita*¹²—would also be rejected. For it too requires individuals to accept the greater gains of others as a sufficient reason for extreme sacrifices.¹³) Now since a principle is just iff it would be chosen in the initial position,¹⁴ a rejected principle is unjust. Therefore, utilitarianism conflicts with justice.¹⁵

Rawls often encapsulates the foregoing argument by means of the slogan that utilitarianism doesn't take seriously the distinction between persons. In particular,

¹⁰ *Ibid.*, 60, 302.

¹¹ *Ibid.*, 14.

¹² Richard B. Brandt ("Some Merits of One Form of Rule-utilitarianism") is one major proponent of this view.

¹³ Rawls's case against both types of utilitarianism is much more elaborate than suggested here: see *A Theory of Justice*, 22-33, 161-6, 175-92.

¹⁴ See *ibid.*, 167 for evidence that Rawls indeed subscribes to this thesis, and not merely to the weaker view that a principle is just if it is chosen in the original position.

¹⁵ It is Rawls's notion of justice with which utilitarianism is said to conflict, and yet in his "Two Concepts of Rules" Rawls *defends* utilitarianism against the charge that it licenses (in fact requires) unjust punishment (Rawls dubs it 'telishment'). In a nutshell, Rawls distinguishes between justifying a practice or an institution (such as games and rituals, trials and parliaments) and justifying a particular action falling under it. He then argues that if one is careful to apply the principle of utility also to the *institution* which is to authorize particular acts of telishment, then there is less danger of it justifying them; and this is because such an institution involves enormous hazards (278-9).

How can Rawls's position in "Two Concepts of Rules" be reconciled with the views that he espouses in *A Theory of Justice* and in "Justice as Fairness"? Quite easily, I think. As Rawls himself argues in "Justice as Fairness" (fn. 4), applying the principle of utility not merely to particular actions but to institutions as well *reduces* the danger of it justifying telishment, but does not *eliminate* it.

utilitarianism offers a reasonable choice principle for *individuals*: it undoubtedly makes sense for an individual to undergo sacrifices now for greater benefits later—for example, to diet in order to improve health or looks—because, since it is the *same person* who is being benefited and harmed, the gain compensates for the loss. Now why not regard what is rational for an individual as right for an association of individuals and adopt the principle of utility for *social* choice? Why not balance satisfactions and dissatisfactions between different individuals, just as an individual balances present and future gains against present and future losses? Because the overall greater good may be achieved by harming some to benefit *others*, so that the gains do *not* make up for the losses (unless there is a social entity that undergoes a sacrifice for its own greater good).¹⁶

Note that the losses and gains in question refer to social and economic advantages as well as the basic rights (liberties). According to Rawls then, persons in the original position would choose principles according to which individuals have basic rights the violation of which can never be justified or compensated for either by greater economic and social advantages *or* by a greater sum of realization of rights. That is to say, maximization of (a) social and economic advantages and/or (b) realization of rights by means of violation of rights is forbidden.¹⁷ (It is precisely in this sense that Rawls's theory of justice is deontological.¹⁸) Persons in the original position would therefore reject both (i) a principle that requires enslaving or killing a few for maximum economic

¹⁶ *A Theory of Justice*, 22-7.

¹⁷ An exegetical discussion of (b) appears in Appendix I. The need for such discussion arises because, while being explicit with respect to social and economic advantages, Rawls is regrettably not as forthcoming regarding rights. (Who was it who said that one learns the most about the virtues of a theory by observing its omissions?)

¹⁸ *A Theory of Justice*, 30.

prosperity, and (ii) a principle that requires killing a few to save many (as well as a host of other principles).

Two comments. First of all, in connection with (ii) it's important to recognize that whether the many are threatened by persons (say, an outraged mob) or nature (illnesses, landslide and so on) should make no difference from Rawls's point of view. The reason? Utilitarianism—with its insistence on happiness maximization (or pain minimization)—requires agents to kill a few not only to prevent many killings, but also to prevent many deaths from natural causes. If Rawls's theory of justice forbids only killings to prevent killings, it constitutes a seriously incomplete reply to utilitarianism.

The many then may be threatened by humans or nature. Together with (b) this implies that each of the many has a right to life. Therefore, Rawls must submit that the right to life means not simply the right not to be killed by others, but rather the right not to die, which—unlike the former—corresponds to both negative and positive duties on the part of others (the former corresponds to a negative duty alone).¹⁹

Whether Rawls *in fact* holds that the right to life means the right not to die is hard to tell. On the one hand he argues that the distinction between positive and negative liberties (and hence the distinction between positive and negative duties) is immaterial for his theory.²⁰ For his part, liberty can always be completely explained by a reference to three elements: the agents who are free, the restrictions or limitations which they are free from, and what it is that they are free to do or not to do.²¹ On the other hand he mentions

¹⁹ Negative and positive duties are duties to not interfere and interfere, respectively.

²⁰ *A Theory of Justice*, 201-2.

²¹ *Ibid.*, 202.

no case where the fact that a natural cause constrains a person's right to life corresponds to a positive duty on the part of others: his discussion is limited to the (negative) duty of non-interference.²²

The second remark concerns both (i) and (ii). As Rawls notes,²³ utilitarians *are* in a position to agree that persons have rights. Indeed, Mill defended rights enthusiastically, arguing that a society which did not grant them to its members would be insecure and fearful, and thus unhappy.²⁴ However, the utilitarian can defend rights only as long as they promote utility: he must agree that rights may be violated in case they do not maximize utility.

4. Killing a Few to Save Many (1)

In Rawls's view then, persons in the original position would reject both (i) a principle that requires enslaving or killing a few for maximum economic prosperity, and (ii) a principle that requires killing a few to save many.

I agree that the parties in the initial position are unlikely to select a principle according to which a few should undergo major sacrifices for a greater total amount of *social and economic* advantages. Are they also unlikely to accept a principle that requires killing a few to save many? I think not. *Rational, self-interested* individuals are *likely* to accept such a principle for the simple reason that one's chances to be the victim are much smaller than one's chances to be saved.²⁵

²² *Ibid.*, 202-3.

²³ *Ibid.*, 28.

²⁴ *Utilitarianism*, 189-90, 196.

²⁵ Judith Jarvis Thomson argues similarly in Chapter 6 of her *The Realm of Rights*, especially pp. 183-7, 195, 201.

The question before us, you will recall, is this. Is the tension between utilitarianism and justice over the issue of whether killing a few to save many is unjust—is this tension a matter of degree?

I just argued that it's rational to accept a principle which requires killing a few to save many, because it maximizes survival chances. Now if this is so, then it's equally rational to consent to a principle that requires killing one to save two: for *this* principle too maximizes survival chances.

It follows that justice requires any principle that maximizes survival chances. *But so does utilitarianism!* And if a principle fails to maximize survival chances, neither justice nor utilitarianism require it. Either way then, utilitarianism and justice are in a harmonious relation. And the tension between them is therefore *not* a matter of degree.

5. Killing a Few to Save Many (2)

Is the foregoing reasoning sound? The force propelling it is no doubt the supposition that it's rational to choose a principle that requires killing some to save others *as long as* the number of beneficiaries exceeds the number of victims. This supposition however is false. To see why note that doing harm is worse than merely allowing it (or, what is the same thing, worse than refraining from generating benefit). Therefore, killing is worse than letting die or failing to prevent a death. And therefore again, sufficiently many lives must be saved in order to justify the killing of a single innocent man. *How* many lives is sufficiently many is however unclear: we may expect a *range* of replies to this question.

The upshot of these observations is that sacrificing one to save *very* few others is definitely unjust, sacrificing one to save a few or many others is neither definitely unjust

nor definitely just, and sacrificing one to save *very* many others is definitely just. That is, as the difference between the number of beneficiaries and the number of victims increases, the tension between utilitarianism and justice decreases. Ergo, the clash between these positions *is* after all a matter of degree.

6. The Doctrine of Doing and Allowing: A Thomson-inspired Argument

But why would rational, self-interested persons behind a veil of ignorance subscribe to the doctrine that doing harm is worse than merely allowing it? Why accept a principle on which doing harm is worse than failing to prevent it, a principle on which the (negative) duty to refrain from harming others is more stringent than the (positive) duty to help others in need? Why should we go to almost any length—endure almost any sacrifice—to avoid killing someone but not to save his life? The reason must be that the doctrine of doing and allowing somehow enhances self-interest.²⁶

There appear to be no conclusive arguments to that effect.²⁷ There are however a couple of *inconclusive* ones.

The first argument draws on—but is not identical to—an argument that Judith Jarvis Thomson advances in her *Realm of Rights*.²⁸

Rejecting the doctrine of doing and allowing amounts in effect to embracing either the view that killing and failing to prevent a death are *equally* wrong, or the view that failing to prevent a death is *more* wrong than killing. In what follows I shall

²⁶ It's far from clear that the distinction between doing harm and failing to prevent it can be drawn satisfactorily (see Kagan, *The Limits of Morality*, 101-6). But I shall assume that it can.

²⁷ In point of fact, the doctrine of doing and allowing appears to enjoy not only no conclusive support of this particular kind: it appears to enjoy *no kind* of conclusive support. I briefly elaborate on this remark in Appendix 2.

²⁸ 183-4.

concentrate on the former position, since—being the weaker of the two—whatever is true about it also applies (*mutatis mutandis*) to the latter.

The former position presumably implies (1) that agents may kill one to save another, and (2) that agents may or even should kill one to save *several* others. Would rational, self-interested persons behind a veil of ignorance consent to these principles? (1) keeps an individual's chances of survival constant, since being saved and being victimized are equally probable. (2) on the other hand actually *increases* a person's probability of survival, as one is likelier to be saved than victimized. It seems therefore that the doctrine of doing and allowing not only doesn't enhance—but actually *undermines*—self-interest.

Does it really? As things currently stand in moderately affluent or affluent societies, the principles at issue are applicable primarily to cases of organ shortage. (1) licenses killing a person for his organs to save another; (2) licenses or requires killing one for his organs to save several others. Should the parties in the original position consent to *these* principles? Again, while (1) leaves a person's probability of survival unaltered, (2) actually increases it. (Imagine, to take a concrete example, that society adopted a specific version of (2): a healthy individual should be cut up for his organs whenever five others are in need of parts. Since an individual is five times likelier to be saved than cut up, consenting to this principle increases one's chances of survival.) It seems therefore that the parties in the original position should consent to at least (2).

Should they really? Note that—given human psychology²⁹—principles (1) and (2) would in all likelihood encourage lifestyles and conditions that tend to diminish life expectancy. In particular, highly demanding yet health-promoting conduct such as proper diet and exercise would be given up in favor of damaging yet tempting habits such as smoking and drinking. This would be so not merely because illness becomes less threatening once organs are easily replaceable, but also because being in good health would become undesirable as it would make victimization likelier. Now as the incidence of illness increases, so does the demand for parts and hence the number of individuals being cut up for them. The healthy would thus comprise an ever shrinking portion of the populace. Consequently, there would, at a certain point, be much demand for—and no supply of—organs, a state of affairs inevitably leading to high death rates.

Moreover, since it is the healthy who are being cut up to save the ill, and since a new (say) heart or kidney does not cure a genetic propensity for heart or kidney failure, hereditary ailments would become more prevalent: which in turn would raise the demand for parts. Finally, adoption of principles (1) and (2) would discourage medical research, again a state of affairs leading to higher death rates.

The parties in the original position should therefore *reject* principles (1) and (2): which means that far from undermining self-interest, the doctrine of doing and allowing is in fact conducive to it. (As Jarvis Thomson remarks, the doctrine of doing and allowing does undermine self-interest in a world in which there is nothing whatever that can be done to affect one's health (ailments that issue in organ failure simply descend on

²⁹ The parties in the original position need not be ignorant about *general* facts concerning human psychology and human society; quite the contrary, they are presumed to know whatever general facts affect the choice of the principles of justice (*A Theory of Justice*, 137-8).

people by chance). In *such* a world, rejecting the doctrine *would* increase one's probability of survival.³⁰ This comment of Jarvis Thomson's highlights why the argument at hand is inconclusive.)

The foregoing argument can be cast more explicitly as follows:

1. Rejecting the doctrine of doing and allowing amounts to maintaining that failing to prevent the death of another is as wrong as causing it.
2. Embracing the view that failing to prevent the death of another is as wrong as causing it amounts to accepting a principle that permits (requires) killing a person for his organs in order to save another (others).
3. A principle that permits (requires) killing a person for his organs in order to save another (others) diminishes a person's probability of survival.
4. A principle that diminishes one's chances of survival undermines self-interest.
5. From (1)-(4) it follows that rejecting the doctrine of doing and allowing undermines self-interest.

7. The Doctrine of Doing and Allowing: Williams's Argument

The second argument—Bernard Williams's³¹—says that rejecting the doctrine—holding that letting die is more wrong than or as wrong as killing—will most likely ensue in killings designed to prevent deaths. This being so, an escalation of pre-emptive activity is empirically probable; and the total consequences of this will be worse than if it had never started.

Why is an escalation of pre-emptive activity empirically probable? I know of no answer to this question, let alone a *satisfactory* answer. Thus, while enjoying a fair amount of *prima facie* plausibility, Williams's argument is nonetheless inconclusive.

8. The Doctrine of Doing and Allowing: Conclusion

Though the foregoing arguments are separately inconclusive, together they seem to lend sufficient support for the claim that the doctrine of doing and allowing enhances

³⁰ *The Realm of Rights*, 185.

³¹ *Morality*, 96.

self-interest. Accordingly, the doctrine—that is, the view that killing is worse than letting die—would be accepted in the initial position. This means that sufficiently many lives must be saved in order to justify the killing of a single innocent man. (As noted, *how* many lives is sufficiently many is a matter of controversy.)

Thus, sacrificing one to save *very* few others is definitely unjust, sacrificing one to save a few or many others is neither definitely unjust nor definitely just, and sacrificing one to save *very* many others is definitely just. That is, as the difference between the number of beneficiaries and the number of victims increases, the tension between utilitarianism and justice decreases. Ergo, the clash between these views is a matter of degree.

9. The Bombshell

Things however are not so simple. I said that the doctrine of doing and allowing would be consented to in the original position because it maximizes the survival chances of an individual. Now a principle maximizes the survival chances of an individual *behind a veil of ignorance* when, and only when, it maximizes *everyone's* chances of survival. An example contrived by Philippa Foot demonstrates this point effectively.³²

Suppose you are the driver of a trolley. The trolley rounds a bend, and there come into view ahead five track workmen, who have been repairing the track. The track goes through a bit of a valley at that point, and the sides are steep, so that you must stop the trolley if you are to avoid running the five men down. You step on the breaks, but alas they don't work. Now you suddenly see a spur of track leading off to the right. You can turn the trolley onto it, and thus save the five men on the track straight ahead. Unfortunately, Mrs. Foot has arranged that there is one track workman on that spur of track. He can no more get

³² "The Problem of Abortion and the Doctrine of Double Effect."

off the track in time than the five can, so you will kill him
if you turn the trolley onto him.³³

If the workmen are assigned to the tracks randomly, then prior to knowing who will end up working on the right track, they would—if they are rational—agree that you should turn the trolley to the right. After all, such conduct on your part maximizes their chances of survival.³⁴ ³⁵ Yet such conduct on your part also maximizes survival chances *in general*: which means that—as far as survival is concerned—*the interest of a person behind a veil of ignorance converges with the general interest*.³⁶

Two facts explain this result. The first is that survival is an all-or-nothing (binary) property: one either completely survives or one doesn't. Since the distribution of such a property cannot be but equal, maximizing survival chances in general maximizes the number of individual survivors. (Or—as Rawls would have it—as far as binary properties are concerned, the distinction between persons is taken care of naturally.) The second fact is that the veil of ignorance ensures that no one is in a position to design principles that would favor his own condition: which means that in order to maximize her own chances of survival, a person behind a veil of ignorance must maximize survival

³³ The description is Judith Jarvis Thomson's ("The Trolley Problem")

³⁴ See also Jarvis Thomson, *The Realm of Rights*, 181-2.

³⁵ What about principles that merely increase—without maximizing—one's chances of survival? Well, consider a variation of the trolley problem, where there is an additional spur of track leading off to the left, and three men are working on it. Under such circumstances, turning to the right *maximizes* everyone's probability of survival, whereas turning to the left merely *increases* this probability. Since the workers are rational, they would all agree that you should turn the trolley to the right. And they would also agree that if turning to the right becomes impossible for some reason, you should turn to the left (and not continue straight ahead). However, if turning to the right is impossible, turning to the left does not merely increase everyone's chances of survival; it *maximizes* these chances. This shows that principles which merely increase, without maximizing, everyone's probability of survival are of little relevance to the present discussion.

³⁶ An individual's interest in her own survival is normally unaffected by the fact that all other persons are also interested in their own survival. Note however that some preferences of individuals do depend on the preferences of others. Suppose that of several alternatives available to me, I prefer lying on the beach the most. Knowing that millions of others have the same preference is likely to affect mine.

chances in general. Taken together, these two facts imply that by maximizing survival chances in general, a person behind a veil of ignorance willy nilly maximizes his own chances of survival, and vice versa.³⁷

Now since a principle maximizes the survival chances of an individual behind a veil of ignorance when, and only when, it maximizes survival chances in general, if the doctrine of doing and allowing maximizes an individual's chances of survival, it also maximizes everyone's chances of survival. But as such, it would be *embraced* by the utilitarian; and she too would adopt the view that sacrificing one to save *very* few others is unjust. This means that, as far as survival is concerned, utilitarian requirements never conflict with those of justice. The tension between these views is *not* a matter of degree.

Here is the same point seen from a slightly different angle. Since a principle maximizes the survival chances of an individual behind a veil of ignorance when, and only when, it maximizes survival chances in general, consenting to a principle which maximizes the survival chances of an individual entails consenting to a principle that maximizes survival chances in general; which in turn means deeming the latter just. But the latter is a form of utilitarianism. Therefore, this sort of utilitarianism is just: it's *not at all* incompatible with justice. The disharmony between the views is *not* a matter of degree.

10. Who Should the Victim Be?

Another observation strengthens this conclusion. Consider a life boat type of situation, where one person must be sacrificed to allow for the survival of some others. Suppose, to fill in some detail, that after days in a life boat, a group of five individuals

³⁷ This conclusion is reminiscent of Hare's. For Hare maintains that Rawls's theory, in its simplest and most natural form, leads to utilitarianism of a sort ("Rawls' Theory of Justice," 91, 94, 101-4).

finally lands on an uninhabited island. The five include: Avrum, male, a forty-seven year old mechanical engineer; Hinda, female, a twenty-seven year old multi-media artist; Shoshana, female, a fifty-six year old doctor; Yidl, male, a gifted eleven year old violinist; and Yankel, male, a thirty-three year old graduate student of philosophy. One of these people must be killed, if they are to survive. Who should it be? Deciding the issue by lot would be very unwise, for the result may be Shoshana or Avrum, whose knowledge is essential for the survival of all. Or it might be Hinda, the only woman of child bearing age. In short, it must be the least useful person, presumably Yidl or Yankel. Utilitarian considerations run deep.

The classical utilitarians denied that some people are less valuable than others. Mill attributes to Bentham the dictum 'everybody to count for one, nobody for more than one,'³⁸ which—according to Mill—follows from the principle of utility.³⁹ The derivation he had in mind seems to be the following:

1. One should always maximize total happiness (this is the principle of utility).
2. The total amount of happiness consists in the collection of happiness experienced by individuals.
3. All are equally capable of experiencing happiness.
4. If some persons count for more than one, a violation of the principle of utility is likely to ensue. To take an example, suppose that *A*, *B*, and *C* experience one unit of happiness each. If *A* and *B* count for one each, and *C* counts for three, the happiness of *A* and *B* would—should the need arise—be sacrificed in favor of that of *C*. Given that all are equally capable of experiencing happiness, such conduct results in one unit of happiness. If however all count for one, *C*'s happiness would—should the need arise again—be sacrificed for that of *A* and *B*; for such conduct results in *two* units of happiness. Thus, if some count for more than one, agents will often fail to maximize happiness, thereby violating the principle of utility.
5. Therefore, the principle of utility requires that one person's happiness is counted for exactly as much as another's.

³⁸ *Utilitarianism*, 199.

³⁹ *Ibid.*, 198.

Well, it may be true that all are equally capable of experiencing happiness. Yet it's doubtful that all are equally capable of *bringing* happiness. Being a doctor, Shoshana is more capable of bringing—or capable of bringing more—happiness than a graduate student of philosophy; which means that she should count for more than him. The precept 'everybody to count for one' in fact *conflicts* with the principle of utility. (If, as I'm inclined to think, Mill's argument is sound, it follows that both rejecting and endorsing this precept result in violations of the principle of utility. This problem can be solved by refining the precept: it should state in particular that all are equal as far as the capability for *experiencing* happiness is concerned; but as far as the capability for *bringing* happiness is concerned, some should count for more than one.)

Should rational, self-interested persons behind a veil of ignorance (i.e., not knowing whether they are the least useful individual in a certain community or group) accept a principle that requires killing the least useful person in order to save others? Hard to tell. To the extent that a doctor improves everybody's chances of survival, accepting the principle would be rational and hence just. Yet there may be subtler reasons for thinking that this principle doesn't in fact maximize one's probability of survival. In that case however, utilitarianism doesn't prescribe it. Either way then, utilitarianism and deontological justice make compatible demands: they are not at all incompatible. The tension between them is *not* a matter of degree.

11. What is a Matter of Degree

However, the moral issue at hand (whether killing a few to save many is unjust) is a matter of degree. For according to both utilitarianism and justice, the rightness degree of killing an innocent person gradually changes—from wrong to right or vice versa—as a

function of a scalable property: namely, the difference between the number of beneficiaries and the number of victims. (Provided of course that the doctrine of doing and allowing indeed maximizes an individual's probability of survival.)

It may be thought that this difference is the relevant scalable property only *loosely* speaking; that *strictly* speaking, the property is in fact a product of that difference and the values of the individuals in question. Why might one think that? Because—as I argued in the previous Section—certain individuals, such as Shoshana or Avrum, are more valuable than others under certain conditions. And if this is so, then saving the lives of a relatively small number of valuable individuals justifies the killing of an innocent person, especially if the latter is not as valuable. (Indeed, it may very well turn out that saving the life of an incredibly valuable person justifies the killing of several “worthless” people.) On the other hand, if the would-be beneficiaries are not as fortunate (they fail to be particularly valuable), only a relatively large number of them would justify the killing of an innocent man, especially if the latter is very valuable.

Recall however that what ultimately makes individuals valuable is their greater capability for saving other lives. So that by saving Shoshana's life one also saves (indirectly) the lives that *she* would later save. This being so, sacrificing (say) three “worthless” people to save Shoshana wouldn't—under certain circumstances—amount to saving *one* by killing three: it amounts rather to saving *many* by killing three. The moral issue in question then is *strictly* speaking a function of the difference between the number of beneficiaries (direct *and indirect*) and the number of victims. Nothing more.

12. Enslaving a Few to Save Many

So far I have considered trade-offs of only lives for lives. Let's now consider briefly whether a few should be enslaved to prevent (how many?) deaths, or—more generally—whether one kind of freedom may be exchanged for another. Since presumably some freedoms are more weighty than others (e.g., the right to life is more weighty than the right to the products of one's own labor), persons in the original position would in all likelihood agree to principles that require exchanges of only lower- for higher-ranked rights. They would thus accept a principle which requires (say) enslaving a few to prevent a few (and *a-fortiori* many) deaths, but would reject a principle that requires killing a few to prevent the enslavement of many (and *a-fortiori* of a few).

What is the rule governing such exchanges? To simplify the inquiry into this question, suppose that in a certain society only three rights matter—the right to life, the right to ownership and control of one's own body (freedom of the person) and therefore to the products of one's own labor, and the right to freedom of speech—with the first mattering more than the second and the second more than the third. Exchanging the freedom of speech of many for the right to life of an innocent man (should the need arise) seems acceptable, as does exchanging the freedom of the person of a few for the right to life of a single innocent man.

Two considerations are involved then. The first concerns the difference in importance between waived and gained rights, the second—the difference in numbers between victims and beneficiaries. The bigger the first (second), the smaller the second (first) needs to be. And the bigger (smaller) the differences, the weaker (stronger) our resistance to rights violation. For some examples, depriving two persons of their right to

free speech in order to save hundreds of lives generates little or no resistance: indeed we may feel an *obligation* to violate the rights of the two. Enslaving millions to save one life creates much resistance: we may feel that such an act is *forbidden*. And enslaving millions to prevent hundreds of deaths leaves the matter undecided: one has no clear intuition as to whether such acts are obligatory, merely permissible, or rather forbidden.

Would utilitarianism and justice make incompatible demands with respect to at least some of the exchanges in question? As we have seen, as far as binary properties—such as survival—are concerned, the interests of a person behind a veil of ignorance converge with the general interest. So the question before us boils down to another one: that is, whether—like the right to life—freedom of the person and freedom of speech also rest on binary properties. *If* they do (I give up pursuit of this question here), then a certain exchange best serves self-interest when and only when it best serves the general interest. An exchange that best serves self-interest would be consented to by persons in the initial position and is therefore just. And an exchange that best serves the general interest would be required by utilitarianism. *Ergo*, *if* freedom of the person and freedom of speech also rest on binary properties, utilitarianism and justice would require the exact same exchanges: they would never issue in incompatible demands.

13. Summary

Two claims best summarize this Chapter. The first is the conclusion of the following argument:

1. Any principle that would be chosen in the initial position is just by Rawls's standards.
2. Any principle that maximizes an individual's probability of survival would be chosen in the initial position.
3. From (1) and (2) it follows that any principle which maximizes an individual's probability of survival is just by Rawls's standards.

4. A principle maximizes the survival probability of an individual behind a veil of ignorance when and only when it maximizes everyone's probability of survival.
5. From (3) and (4) it follows that any principle which maximizes everyone's probability of survival is just by Rawls's standards.
6. A principle which maximizes everyone's probability of survival is nothing but a form of the principle of utility: a theory consisting of it is nothing but a form of utilitarianism.
7. From (5) and (6) it follows that this type of utilitarianism is just by Rawls's standards: there is no conflict whatsoever between it and Rawls's notion of justice.
8. From (7) it follows that the conflict between justice and the type of utilitarianism under consideration is not a matter of degree.

The second claim is that if the doctrine of doing and allowing enhances self-interest, then the issue of whether killing a few to save many is unjust *is* a matter of degree. The rightness degree of killing an innocent changes as a function of a scalable property: namely, the difference between the number of beneficiaries and the number of victims. As the difference increases, the wrongness degree involved in killing an innocent decreases.

Appendix 1: Rawls on Maximization of Rights Realization

Rawls believes that individuals in the initial position would not trade rights for a greater sum of *social and economic* advantages. Does he also believe that they wouldn't exchange rights for a greater sum of realization of *rights*? Does he, that is to say, deny "utilitarianism of rights" (Nozick's terminology⁴⁰; I shall henceforth drop the quotes)—the view that the rights of some may be transgressed in order to maximize (minimize) their overall realization (transgression). If he doesn't, it means that he doesn't think that killing one to save many is forbidden.

At times Rawls does appear to be advancing utilitarianism of rights. Why think that? He maintains that the equal liberties may be restricted (only) to avoid an even greater loss of liberty. Such restrictions can apply to either extent or equality. As to the former, if the liberties are left unrestricted they collide with each other; therefore, a less extensive liberty strengthens the total system of liberty shared by all.⁴¹ For example (Rawls's own), rules of order are necessary for intelligent and profitable discussion. Such rules limit our freedom of speech, since we cannot speak whenever we please, but they are required to gain the benefits of this liberty.⁴² Turning to equality, restrictions on it must be acceptable to those with the lesser liberty. For example, the precept one person one vote may be violated on the ground that persons with greater intelligence and education should have extra votes and thus greater influence, since such an arrangement

⁴⁰ *Anarchy, State and Utopia*, 28.

⁴¹ *A Theory of Justice*, 302.

⁴² *Ibid.*, 203.

is in everyone's best interest (all benefit when the opinions of the wiser and more knowledgeable have superior weight).⁴³ Now Rawls's position on the latter type of restrictions seems to imply not only that the right to life of a few may be sacrificed for the right to life of many, but also that a few may be enslaved to prevent the enslavement of many or to prevent a few deaths and so on.⁴⁴

As a matter of fact however, Rawls maintains that violations of equality are acceptable provided that they benefit those with the lesser liberty *themselves*. This proviso—nothing but a corollary of the accusation that utilitarianism does not take seriously the separateness of persons—clearly excludes killing, enslaving, or indeed even restricting the liberty of conscience of some to save *others*. Moreover, in connection with tolerance Rawls explicitly rejects rights utilitarianism, claiming that

...even when the freedom of the intolerant is limited to safeguard a just constitution, this is not done in the name of maximizing liberty. The liberties of some are not suppressed simply to make possible a greater liberty for others. Justice forbids this kind of reasoning in connection with liberty as much as it does in regard to the sum of advantages.⁴⁵

In Rawls's view then persons in the original position would choose principles according to which individuals have basic rights the violation of which can never be justified or compensated for either by greater economic and social advantages *or* by a greater sum of realization of rights.

⁴³ *Ibid.*, 232-3. This position—which Rawls uses to illustrate his point—was actually advanced by J. S. Mill.

⁴⁴ Support for this deduction may be found in Section 7 of Rawls's "Justice as Fairness," where he writes
If one examines the circumstances in which a certain tolerance of slavery is justified, or perhaps better, excused, it turns out that these are of a rather special sort. Perhaps slavery exists as an inheritance from the past and it proves necessary to dismantle it piece by piece; at times slavery may conceivably be an advance on previous institutions (347).

⁴⁵ *A Theory of Justice*, 220.

Appendix 2: The Doctrine of Doing and Allowing

Although defending deontological constraints requires in turn a defense of the doctrine of doing and allowing, few champions of deontology have undertaken that task. Among those who have, one usually witnesses either of two strategies. The first strategy is blatantly circular. It involves pointing out that if the doctrine of doing and allowing were rejected, then some unacceptable result would ensue: yet the reason why the result is unacceptable presupposes the very doctrine that the argument is designed to support. One example was given in Chapter 2, Section 8. Here are two further examples:

1. If the doctrine of doing and allowing is incorrect, then it should not be of special importance to people that they not maltreat others *themselves*. I.e., refraining from maltreating Schwartz themselves and preventing Weiss from harming Schwartz should be of equal importance to people.
2. Yet it *is* of special importance to people that they not maltreat others themselves.
3. Therefore, the doctrine of doing and allowing is correct.

Why is premise (2) correct? Because maltreating others is morally worse than failing to prevent harm to them: presumably, if maltreating others and failing to prevent harm to them were equally bad from a moral point of view, they would be equally important to people.

1. If the doctrine of doing and allowing is incorrect, then if killing one to save five is impermissible so is just passing by a dying person on the way to saving the five.
2. Surely however killing one to save five is impermissible, while just passing by a dying person on the way to saving the five is permissible or even obligatory.
3. Therefore, the doctrine of doing and allowing is correct.⁴⁶

Why is (2) true? Because killing is worse than letting die.

⁴⁶ This argument is Frances Kamm's ("Harming Some to Save Others," 232).

The second strategy for defending the doctrine of doing and allowing consists in invoking a further moral principle.⁴⁷ But how could a principle *explain* a principle at the same level of analysis?

In light of this lack of adequate support, some have concluded that the doctrine of doing and allowing is highly questionable.⁴⁸

⁴⁷ See for example Kamm, *Non-consequentialism, the Person as an End-in-itself and the Significance of Status*, 368-9. Or her "Harming Some to Save Others," 243, 254-5.

⁴⁸ E.g., Kagan, *The Limits of Morality*, 115-21; Scheffler, *The Rejection of Consequentialism*, 15, 103.

Bibliography

- Altman, Andrew, Michael Bradie and Fred D. Miller. "On Doing Without Events." *Philosophical Studies* 36 (1979): 301-7.
- Anscombe, G. E. M. *Intention*. Oxford: Blackwell, 1957.
- , "Modern Moral Philosophy." 1958. In Cahn and Haber.
- , "Under a Description." *Nous* 13 (1979): 219-233.
- Aristotle. *Nicomachean Ethics*. In *Introduction to Aristotle*. Ed. Richard McKeon. New York: Random House, 1947.
- Atwell, John E. "The Accordion Effect Thesis." *Philosophical Quarterly* 19 (1969): 337-42.
- Bach, Kent. "Actions are not Events." *Mind* 89 (1980): 114-20.
- Barcan Marcus, Ruth. "Moral Dilemmas and Consistency." 1980. In Cahn and Haber.
- Beckman, Erik. *The Criminal Justice Dictionary*. 2d ed. Ann Arbor: Pierian Press, 1983.
- Bennett, Jonathan. *Events and their Names*. Oxford: Clarendon Press, 1988.
- , "Shooting, Killing and Dying." *Canadian Journal of Philosophy* 2 (1973): 315-23.
- Bentham, Jeremy. *Introduction to the Principles of Morals and Legislation*. New York: Hafner, 1948.
- Brandt, Richard B. "Some Merits of One Form of Rule-utilitarianism." *University of Colorado Studies* (1967): 39-65.
- Cahn, Steven M., and Joram G. Haber. *Twentieth Century Ethical Theory*. Englewood Cliffs: Prentice Hall, 1995.
- Campbell, C. A. *Selfhood and Godhood*. London: Allen & Unwin, 1957.
- Casati, Roberto, and Achille Varzi. *Events*. Brookfield: Dartmouth, 1996.
- Chisholm, Roderick. "Supererogation and Offence: A Conceptual Scheme for Ethics." *Ratio* 5 (1963): 1-14.
- Chisholm, Roderick, and Ernest Sosa. "Intrinsic Preferability and the Problem of Supererogation." *Synthese* 16 (1966): 321-31.

- , "On the Logic of 'Intrinsically Better'." *American Philosophical Quarterly* 3 (1966): 244-9.
- Cooper, John. *Reason and Human Good in Aristotle*. Cambridge, MA: Harvard UP, 1975.
- Danto, Arthur. 1965. "Basic Actions." *American Philosophical Quarterly* 2 (1965): 141-8.
- , "What We Can Do." *Journal of Philosophy* 60 (1963): 435-45.
- Davidson, Donald. 1963. "Actions, Reasons and Causes." In Davidson, *Essays*.
- , "Agency." 1971. In Davidson, *Essays*.
- , *Essays on Actions and Events*. Oxford: Clarendon Press, 1980.
- , "Events as Particulars." 1970. In Davidson, *Essays*.
- , "The Individuation of Events." 1969. In Davidson, *Essays*.
- , "The Logical Form of Action Sentences." 1967. In Davidson, *Essays*.
- Davis, Lawrence. "The Individuation of Actions." *Journal of Philosophy* 67 (1970): 520-30.
- Dreier, James. "Structures of Normative Theories." *The Monist* 17 (1993): 22-40.
- Driver, Julia. "The Suberogatory." *Australasian Journal of Philosophy* 70 (1992): 288-95.
- Dworkin, Ronald. *Taking Rights Seriously*. Cambridge MA: Harvard UP, 1977.
- Edgington, Dorothy. "On Conditionals." *Mind* 104 (1995): 235-329.
- Feinberg, Joel. "Action and Responsibility." *Philosophy in America*. Ed. Max Black. Ithaca: Cornell UP, 1965.
- Foot, Philippa. "The Problem of Abortion and the Doctrine of Double Effect." *Oxford Review* 5 (1967): 5-15.
- , "Virtues and Vices." 1978. In Cahn and Haber.
- Forbes, Greame. "Thisness and Vagueness." *Synthese* 54 (1983): 235-59.
- Fried, Charles. *An Anatomy of Values*. Cambridge MA: Harvard UP, 1971.

- Gauthier, David. *Morals by Agreement*. Oxford: Clarendon Press. 1986.
- Goldman, Alvin. *A Theory of Human Action*. Princeton: Princeton UP. 1976.
- Goodman, Nelson. *The Structure of Appearance*. Cambridge MA: Harvard UP. 1951.
- Gowans, Christopher W, ed. *Moral Dilemmas*. New York: Oxford UP. 1987.
- Hare, R. M. "Rawls' Theory of Justice." In *Reading Rawls*. 2d ed. Ed. Norman Daniels. Stanford: Stanford UP. 1973.
- Henson, Richard. "What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action." *Philosophical Review* 88 (1979): 39-54.
- Herman, Barbara. *The Practice of Moral Judgment*. Cambridge MA: Harvard UP. 1993.
- Heyd, David. *Supererogation*. Cambridge: Cambridge UP. 1982.
- Horgan, Terence. "The Case Against Events." *The Philosophical Review* 87 (1978): 28-47.
- Hornsby, Jennifer. *Actions*. London: Routledge & Kegan Paul. 1980.
- Jarvis Thomson, Judith. "A Defense of Abortion." *Philosophy and Public Affairs* 1 (1971): 47-66.
- , "Individuating Actions." *The Journal of Philosophy* 68 (1970): 774-81.
- , *The Realm of Rights*. Cambridge MA: Harvard UP. 1990.
- , "The Time of a Killing." *The Journal of Philosophy* 68 (1971):115-32.
- , "The Trolley Problem." In *Rights, Restitution, and Risk*. Ed. William Parent. Cambridge MA: Harvard UP. 1985.
- Kagan, Shelly. *The Limits of Morality*. Oxford: Clarendon Press. 1989.
- Kamm, Frances. "Harming Some to Save Others." *Philosophical Studies* 57 (1989): 227-60.
- , "Non-consequentialism, the Person as an End-in-itself, and the Significance of Status." *Philosophy and Public Affairs* 21 (1992): 354-389.
- , "Supererogation and Obligation." *The Journal of Philosophy* 82 (1985): 118-38

- Kant, Immanuel. *Groundwork of the Metaphysic of Morals*. Trans. H. J. Paton. New York: Harper & Row. 1964.
- , *Metaphysic of Morals*. Trans. Mary Gregor. Cambridge: Cambridge UP. 1996.
- Kenny, Anthony. *Action, Emotion and Will*. London: Routledge & Kegan Paul, 1963.
- Kim, Jaegwon. "Events as Property Exemplifications." In *Action Theory*. Eds. Myles Brand and Douglas Walton. Dordrecht: Reidel, 1980.
- Leonard, Henry, and Nelson Goodman. "The Calculus of Individuals and Its Uses." *Journal of Symbolic Logic* 5 (1940): 45-54.
- Lewis, David. "How to Define Theoretical Terms." *The Journal of Philosophy* 67 (1970): 427-46.
- , *The Plurality of Worlds*. Oxford: Blackwell, 1986.
- Lukasiewicz, Jan, and Alfred Tarski. "Investigations into the Sentential Calculus." In *Logic, Semantics, Metamathematics*. Ed. Alfred Tarski. Oxford: Clarendon Press, 1956.
- McCloskey, H. J. "An Examination of Restricted Utilitarianism." *The Philosophical Review* 66 (1957): 466-85.
- , "A Non-utilitarian Approach to Punishment." In *Contemporary Utilitarianism*. Ed. Michael D. Bayles. Garden City: Doubleday, 1968.
- Mill, J. S. *Utilitarianism*. In *On Liberty and Other Essays*. Ed. John Gray. Oxford: Oxford UP, 1991.
- Murphy, Jeffrie G., and Jules L. Coleman. *Philosophy of Law*. Boulder: Westview Press, 1990.
- Nagel, Thomas. "Autonomy and Deontology." In *Consequentialism and its Critics*. Ed. Samuel Scheffler. Oxford: Oxford UP, 1988.
- , "The Fragmentation of Value." 1979. In Gowans.
- , "Moral Luck." 1976. In Chan and Haber.
- , "War and Massacre." In *Mortal Questions*. Cambridge: Cambridge UP, 1979.
- Nielsen, Kai. "Against Moral Conservatism." *Ethics* 82 (1972) :113-24.
- Norcross, Alastair. "The Scalar Approach to Morality." Forthcoming.

- Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books. 1974.
- , "Moral Complications and Moral Structures." 1968. In Nozick, *Socratic Puzzles*. Cambridge MA: Harvard UP. 1997.
- Parsons, Terence. "Some Problems Concerning the Logic of Grammatical Modifiers" *Synthese* 21 (1970): 320-33.
- Paton, H. J. *The Categorical Imperative*. London: Hutchinson. 1967.
- Pettit, Philip. "Utilitarianism." In *Three Methods of Ethics*. Oxford: Blackwell. 1997.
- Plato. *Apology*. In *Complete Works*. Ed. John M. Cooper. Indianapolis: Hackett. 1997.
- , *Republic*. In *Complete Works*. Ed. John M. Cooper. Indianapolis: Hackett. 1997.
- Raiiiton, Peter. "Alienation, Consequentialism and the Demands of Morality." *Philosophy and Public Affairs* 13 (1984): 134-71.
- Rawls, John. "Justice as Fairness." 1958. In Cahn and Haber.
- , *Theory of Justice*. Cambridge MA: Belknap Press of Harvard UP. 1971.
- , "Two Concepts of Rules." 1955. In Cahn and Haber.
- Rescher, Nicholas. *Many-valued Logic*. New York: McGraw Hill. 1969.
- Rickless, Samuel. "The Doctrine of Doing and Allowing." *The Philosophical Review* 106 (1997): 555-575
- Ross, W. D. *The Right and the Good*. Oxford: Clarendon Press. 1930.
- Sartre, Jean Paul. "Existentialism and Humanism." In *The Existentialist Tradition*. Ed. Langulli N. New York: Anchor. 1971.
- Scheffler, Samuel. *The Rejection of Consequentialism*. Oxford: Clarendon Press. 1982.
- Sen, Amartya. "Utilitarianism and Welfarism." *Journal of Philosophy* 76 (1979): 463-89.
- Simmons, Keith. "Kant on Moral Worth." *History of Philosophy Quarterly* 6 (1989): 85-100.
- Singer, Marcus. "Actual Consequence Utilitarianism." *Mind* 86 (1977): 67-77.

- Singer, Peter. "Famine, Affluence, and Morality." *Philosophy and Public Affairs* 1 (1972): 229-43.
- Sinnott-Armstrong, Walter. *Moral Dilemmas*. Oxford: Basil Blackwell, 1988.
- Slote, Michael. *Common-sense Morality and Consequentialism*. London: Routledge & Kegan Paul, 1985.
- , "Virtue Ethics." In *Three Methods of Ethics*. Oxford: Blackwell, 1997.
- Smart, J. J. C. "An Outline of a System of Utilitarian Ethics." In *Utilitarianism: For and Against*. Cambridge: Cambridge UP, 1973.
- Smith Holly. "Varieties of Moral Worth and Moral Credit." *Ethics* 101 (1991): 279-303.
- Smith, Patricia. *Omission and Responsibility*. Forthcoming.
- Sophocles. *Antigone*. Ed. Mark Griffith. Cambridge: Cambridge UP, 1999.
- Swanton, Christine. "Profiles of the Virtues." *Pacific Philosophical Quarterly* 76 (1995): 335-44.
- Thalberg, Irving. "Singling Out Actions, their Properties and Components." *Journals of Philosophy* 68 (1971): 781-7.
- , "When Do Causes Take Effect?" *Mind* 84 (1975): 583-9
- von Wright, Georg Henrik. *The Logic of Preference*. Edinburgh: Edinburgh UP, 1963.
- Walker, Nigel. *Sentencing: Theory, Law, and Practice*. London: Butterworths, 1985.
- Williams, Bernard. "A Critique of Utilitarianism." *Utilitarianism: For and Against*. Cambridge: Cambridge UP, 1973.
- , "Ethical Consistency." 1973. In Gowans.
- , *Ethics and the Limits of Philosophy*. Cambridge MA: Harvard UP, 1985.
- , *Morality*. Cambridge: Cambridge UP, 1993.
- , "Persons, Character and Morality." In *Moral Luck*. New York: Cambridge UP, 1981.
- Williamson, Timothy. *Vagueness*. London: Routledge, 1994.