A **ROUTLEDGE** FREEBOOK

Discover the *Routledge Philosophers* Series

Routledge
Taylor & Francis Group

TABLE OF CONTENTS

DISCOVER THE ROUTLEDGE PHILOSOPHERS SERIES

VISIT WWW.ROUTLEDGE.COM

TO BROWSE FULL RANGE OF TITLES IN THE ROUTLEDGE PHILOSOPHERS SERIES

E-INSPECTION COPIES ARE AVAILABLE FOR ALL TITLES IN THE SERIES

Introduction

The *Routledge Philosophers* series, launched in 2005, is one of the leading examples of its kind. Edited by Brian Leiter (University of Chicago) each volume introduces and assesses the arguments of a major Western philosopher. One of the distinguishing features of the series is its roster of highly respected authors, which includes Jonathan Lear, Christopher Shields, Fred Beiser, Paul Guyer, Samuel Freeman and Don Garrett – several of whose work appears here. In this special Freebook we are delighted to present four chapters from four different volumes in the series, ranging from Aristotle's theories of how to live well to Freud's theories of morality and religion. We have tailored the selection around the general topic of ethics and moral philosophy, as this provides the best and most accessible way of showcasing the range of thinkers covered. However, each volume in the series also examines the contributions each figure has made to a much wider range of philosophical topics, such as metaphysics, epistemology, the nature of mind, religion, politics and art and aesthetics.

One of the notable features of the series is that it encompasses figures who, whilst not themselves philosophers, have nevertheless had significant influence on the development of philosophical ideas and theories of human nature, often radical and controversial. In this freebook we have selected a chapter from Jonathan Lear's volume on Freud, which makes an interesting comparison when placed alongside the excerpts from Aristotle, Hume and Kant.

Looking further ahead we have many more exciting volumes to come, including Elizabeth Anscombe and Hannah Arendt, increasing the coverage of women philosophers; Adam Smith, Karl Marx and Soren Kierkegaard; and W.E.B du Bois, pioneer of African-American philosophy.

We hope these selections provide an exciting taste of the range and quality of the Routledge Philosophers.

Visit our website to view information on the books in full, or to purchase a copy. Links are provided at the beginning of each chapter of this FreeBook. If you have any questions, please contact us.

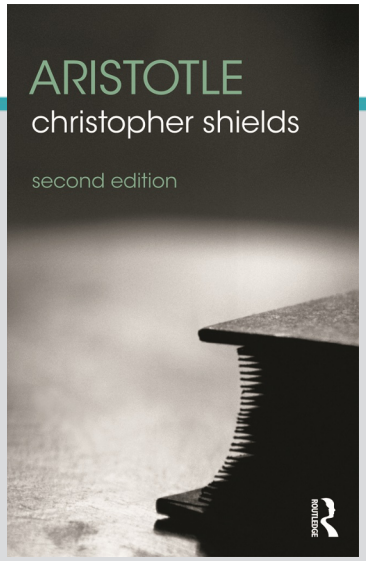
Note to readers: References from the original chapters have not been included in this text. For a fully-referenced version of each chapter, including footnotes, bibliographies, references and endnotes, please see the published title. Links to purchase each specific title can be found on the first page of each chapter.

As you read through this FreeBook, you will notice that some excerpts reference previous chapters – please note that these are references to the original text and not the Freebook.

CHAPTER

1

LIVING WELL



LIVING WELL

Excerpted from *Aristotle*

8.1 The final good for human beings

Human beings engage in purposive behaviour. We do things for reasons and act with ends in view. Thus, we walk to the store *intending* to buy milk. If asked *en route* by a friend whom we meet on the street why we are walking towards the store, the sensible and correct answer is the true one: 'To buy milk'. If our friend is an amusing sort and begins to regale us with jokes and stories so engagingly that we forget where we had been going and why, then we may find ourselves befuddled, forgetting temporarily what we had been doing and trying to recollect the purpose of our being on the street. If we really cannot recall, then we will no longer walk towards the store, for we shall have no purpose motivating us to do so. When we do recall our purpose, then we resume our activity with a sheepish smile upon our face.

Suppose, by contrast, that our friend is not an amusing sort, but a serious-minded philosopher who wants to know *why* we want to buy milk. If we answer again honestly that we want to buy milk for our morning porridge, and she presses on, wanting to know *why* we intend to eat porridge in the morning, then we may well answer that we find porridge healthy and delicious, especially with milk, which we may then excuse ourselves to buy. Insensible of our lack of interest, the philosopher may persist, wanting to know *why* we desire to eat delicious and healthy food. Again, we may respond, that it is because we enjoy delicious food, that eating it brings us pleasure, and that we desire health for the obvious reason that health is good – and, lest it be asked, we all desire good things for ourselves. If we have not by now slipped away, we may hear the philosopher posing the same question, earnestly let us allow, *ad nauseam*, or at least until such time as we say, with exasperation, that we do all these things we do for the sake of happiness. If now asked why we wish to be happy, perhaps rudeness is warranted. We may simply walk away, shrugging and saying that we really must buy our milk.

Although our behaviour is purposive, it seems that such why-questions must leave off at some point. Aristotle finds some significance in these related facets of our behaviour, that we do things for reasons and that our reasons may be subordinated to superordinate reasons until we reach a final and ultimate reason underlying all of our intentional actions. Aristotle opens his *Nicomachean Ethics* with just this commitment, though employing what may seem a disastrous argument on its behalf:

> Every art and every inquiry, and similarly every action and choice, seems to aim at some good; accordingly, the good has rightly been declared to be

LIVING WELL

that at which all things aim.

(EN 1094a1–3)

Unfortunately, even if it is true that there is some ultimate good for all human action, this argument, taken at face value, does not deliver that conclusion. For it may be true that every action aims at some end, even though there is no single end towards which all actions aim. After all, every archer aims at a target, though there is no one target at which all archers aim. If Aristotle argues this way, then he is guilty of the very simple fallacy of noting that everything has some feature and inferring on that basis that there is just one feature had by everything.

That said, it may be possible to understand these lines in a manner more favourable to Aristotle, in either of two ways. First, perhaps he is already assuming in the first line that every intentional action aims ultimately at some one end, the good, and then commenting that it is therefore appropriate that some have characterized the good as that at which all things aim. On this way of understanding these lines, Aristotle does not argue fallaciously, because he does not argue at all. Alternatively, he may be understood as advancing an argument which is not immediately liable to the objection given. Perhaps he means to claim that since every action aims at some sort of end or other, each of which is some sort of good, what these ends have in common is that they are good. Different capitalists market cars, coat hangers, and coffee beans, each aiming for profit in their sector; so profit is rightly called the aim of all capitalists. Similarly, exercise aims at health because health is good, study aims at knowledge because knowledge is good, and recreation aims at relaxation because relaxation is good. What these different sorts of goods have in common is precisely that they are good. Such an inference requires additional work, and may sit uneasily with Aristotle's scruples concerning the univocity of goodness. Even so, it would not implicate Aristotle in the formal fallacy these lines are often taken to commit.

In any case, if we do agree that purposive actions aim at good ends, or at least at apparently good ends, and if we further agree that these ends may be subordinated to one another such that there is some one final good which all humans seek, we would do well to reflect upon the characteristics we expect this final end to have.

To begin, when asked what their final good is, people will likely disagree. Some people, hedonists, will report that what they seek above all else is pleasure. Others, with different priorities, may report that they wish above all to be loved, or that they strive to lead lives of honour, or that riches or power matter most, and so

LIVING WELL

on. Importantly, when they disagree in these ways, parties to this dispute may be disagreeing at either or both of two distinct levels. First, people may agree on the characteristics of the final good, only to wrangle about which states or activities exhibit those characteristics. Or their disagreements may be of a higher order: perhaps these varied answers result from non-equivalent assumptions about what it would take for a state or activity to qualify as a final good. Thus, for example, two people might disagree about relaxation, one suggesting that reading quietly in the library is relaxing while the other recommends water-skiing behind a motor boat as the most relaxing way to spend an afternoon. These people might agree about what relaxation is, but disagree about the best way to achieve it, or they might actually disagree about the nature of relaxation, one supposing that any activity which is non-work related, however vigorous or tiring, qualifies as relaxation, while the other understands relaxation to be restricted to quiet, non-stressful stretches of tranquil inactivity. In order to sort out their disagreement, they would, in the latter case, first need to come to some agreement about the general characteristics of relaxation. In a similar way, those who disagree about the final good for human beings will in some cases first need to reflect on the abstract criteria for something's qualifying as a final good in the first place.

Aristotle's procedure is to begin at this more abstract level. His method recommends that in order to determine the final good, we should first agree about what criteria it must satisfy (*EN* 1094a22–27). Only in this way, he supposes, will substantive agreement paving the way for real progress be possible. Aristotle lays down as conditions for the final good that: (i) it be pursued for its own sake (*EN* 1094a1); (ii) we wish for other things for its sake (*EN* 1094a19); (iii) we do not wish for it on account of other things (*EN* 1094a21); (iv) it be complete (*teleion*), in the sense that it is always choice-worthy and always chosen for itself (*EN* 1097a26–33); and (v) it be self-sufficient (*autarkês*), in the sense that its presence suffices to make a life lacking in nothing (*EN* 1097b6–16). The first three of these conditions are reasonably straightforward, though it is necessary to note that (i) and (iii) are distinct, in that (i) holds that other things are done for its sake, whereas (iii) requires that it not be done for anything beyond itself. One might, for instance, pursue health both for its own sake, because it is an intrinsic good, and also on behalf of something more final than it, because it is regarded as a necessary component of a happy life, with the result that one wants health both for its own sake and for the sake of happiness. On that assumption, health would satisfy (i) but not (iii), and so would fail to be a final end according to the criteria given.

LIVING WELL

The last two criteria are a bit more difficult, since Aristotle is fairly brief in his characterizations of them. For an end to be complete (*teleion*, also sometimes rendered as 'final' or 'perfect'), it must not only be desired for nothing beyond itself, but *always* be such that it is choiceworthy in itself. Aristotle implies that something might be desired for its own sake and for the sake of nothing beyond itself, and yet fail to be complete, because circumstances could alter its status. One way a final end could be impervious to contingencies would be by being fully comprehensive. Thus, if happiness were the final good, then this might be due to the fact that it embraces all possible human goods. Contrast this with pleasure, which might normally be good, desired for itself and for nothing beyond itself, but nonetheless come to compete against other goods, perhaps honour, and so be rendered less choiceworthy in that circumstance. Similarly, an end's qualifying as self-sufficient (*autarkês*) is an extremely demanding criterion. Something is self-sufficient if its presence alone is enough to make a life lacking in nothing. Again, something might be self-sufficient because it is an especially comprehensive good, one embracing all forms of human goodness.

It may seem, given the stringency of these demands, that nothing will emerge as the final good for human beings. After all, what is always choiceworthy for itself and such that all by itself it makes a life lacking in nothing? Looked at from this angle, Aristotle's criteria may seem so austere that they are bound to remain unmet. Looked at another way, though, these demands seem just about right. For they are at this stage only helping to focus debates about the character of the final good. If there is some end which qualifies as the final good, the single all-encompassing human good which we seek in all of our actions, then it really should meet the high standard imposed by these criteria. From this perspective, it is easy to agree with these criteria for the final good, because so far we have not also agreed that anything in fact satisfies them. By the same token, if some end does emerge to satisfy them all, we will have a powerful reason to agree that this good deserves its elevated status.

8.2 The character of human happiness: Preliminary considerations

Perhaps, no matter how stringent Aristotle's criteria may appear, we will nonetheless suppose that there is an obvious candidate for the final good for human beings. This final and ultimate reason for all of our action is simply our own happiness: we all wish to be happy. We wish happiness for its own sake, and

LIVING WELL

not for the sake of anything beyond it; we pursue other goods for the sake of happiness; if we have achieved happiness, genuine happiness, then our lives are complete and lacking in nothing; happiness, by itself, suffices to make our lives good lives (*EN* 1097a30–b8). This is why, in fact, we wish for happiness above all else. Moreover, this is why the question 'Yes, but why do you want to be happy?' is otiose. In the domain of purposive behaviour, why-questions come to an end with happiness.

So much seems unobjectionable.We desire happiness. What is it, though, that we desire? It falls to the philosopher engaged in practical philosophy to address this question. For though we all agree that we seek happiness, it turns out that our agreement obscures important forms of disagreement, because we turn out to disagree about the nature of happiness (*EN* 1095a14–21). When queried, some of us will say that happiness consists in warm and fuzzy self-regard; others suppose that happiness is fame; others power; and many more are confident that happiness is pleasure. Aristotle argues that each of these answers is wrong.

To some modern sensibilities, the suggestion that someone could be wrong about her own happiness seems preposterous on its face. After all, I decide what makes me happy; and I know when I am happy and when I am not. Only I can judge when I am happy, and whenever I do so judge, then I am in fact happy. Surely it does not fall to the philosopher sitting in her university office to decide such matters for me.

On the contrary, counters Aristotle, it falls to the philosopher to determine the nature of happiness, since happiness, like other central ethical concepts, admits of analysis. Two features of his approach help explain why he proceeds on this assumption.

To appreciate his eventual account, it is first of all necessary to understand a central feature of his approach. Aristotle is committed to an *objective* account of happiness. We may contrast two ways of thinking about happiness. Let us say an account of happiness is *subjective* if it presupposes that someone's being happy is at least partially constituted by their regarding themselves as happy. Thus, for instance, on one subjective conception, happiness consists in desire satisfaction, such that to be happy is simply to regard oneself as having had one's desires satisfied. Typically, let us suppose, desire satisfaction eventuates in a feeling of warm, even glowing satisfaction and warm self-regard. Thus, on this subjective conception of happiness, an agent can be expected to know when he is happy and to be authoritative about his own happiness. If he *feels* happy, and regards himself

LIVING WELL

as happy, then he is happy, and otherwise not.

On a subjective conception of happiness, it barely makes sense to imagine someone reporting, 'I thought I was happy, but I was mistaken'. By contrast, an *objective* account of happiness holds that happiness consists in satisfying some criteria which are not determined by an agent's desires or self-conception. To be happy, on the objective conception, requires that a person lead a successful and flourishing life, where, again, the conditions of successful living or flourishing are not up to the agent. It is helpful, in this connection, to think about judgments of happiness from a third-person point of view. One might judge that a neighbour or relative is living well, and is flourishing as a human being, even without knowing too much about her interior life. More importantly, one might readily judge that a friend or loved one is *not* living the best life available to him, might lament that he is careening down a path of self-destruction, say by foolish use of harmful drugs, even though, when asked, the person so judged will report, sincerely, that he feels fine, that he is happy. On the objective conception of happiness, we are in principle entitled, in these sorts of cases, to conclude that the person is wrong about his own self-ascription of happiness. In the same way, we may look back at an earlier period of our own lives and judge correctly that while we thought we were happy, we were mistaken.

Now, it is often noted in this connection that what we have been calling 'happiness' is for these reasons an unfortunate translation of Aristotle's word *eudaimonia*, which might better be rendered as 'flourishing' or 'living well' or 'living successfully'. This point about translation can be easily overblown, however: Aristotle appreciates that people disagree about the nature of *eudaimonia*, that 'the many do not give the same answer as the wise' (*EN* 1095a21–22), because they think 'it is something obvious and manifest' (*EN* 1095a22). This is just to say, however, that people disagree about what happiness is, and that some, who are unreflective, simply assume without warrant that its nature is plain for all to see. From Aristotle's perspective, this easy, unreflective contention should not be ceded without a fight.

What really matters in this discussion is not whether we translate *eudaimonia* as 'happiness' or not, but whether, having agreed to call *whatever satisfies the criteria for the final good* happiness, we can uncover some state or activity up to the task. Aristotle's first contention in this regard is that subjective conceptions of happiness fall down miserably on this score. Sometimes our desires are satisfied, but instead of feeling pleasure or satisfaction, we in fact become

LIVING WELL

perplexed with ourselves, even to the point of self-alienation. A man who desires a yellow sports car more than anything, who sacrifices mightily to obtain it, may wonder, when he has it in his possession, why exactly he had wanted it so very badly.

Further, even if we do feel satisfied when our desires are realized, we may in truth do so despite the fact of our having developed desires that are not worthy of us. This point is less obvious, but again we may better appreciate Aristotle's contention by adopting a third-person point of view. A woman might regard her dear son with concern, because he is not living up to his potential. She knows in an unbiased way that he is highly intelligent, exceptionally talented, and superior in his natural athletic abilities. Yet she also sees that he is so eager to impress his deadbeat friends that he is purposely performing poorly, because of his burning need to be accepted. Such a mother will rightly judge that her son is not flourishing, that he is not living the rich life he might. If he regards her as meddlesome in her attention, and informs her that he is happy and just wants to be left alone, then this may simply reflect his immaturity, that he is not really in a position to judge his own circumstance correctly; he has, among other shortcomings, wilful blind spots. If someone now wants to insist that the teenager is nonetheless *happy*, then it need only be pointed out that he is not in any condition which satisfies the criteria for the final good we have accepted. Again, there is no point in squabbling about whether we should render *eudaimonia* as happiness. What matters regarding the case in question concerns whether the boy is leading the best life available to him, whether what he is calling happiness in fact satisfies the criteria for the ultimate human good we have accepted.

Indeed, Aristotle urges, we can see that some common conceptions of happiness do not meet these criteria, and so are to be set aside. One obvious loser is the life of the money-maker (*EN* 1096a6–11). In advancing this argument, Aristotle does not disparage money as such, but observes, correctly, that money is an instrumental good. If it is merely an instrument, then money is not choiceworthy in itself and thus violates the very first of our accepted criteria, namely that the final good be chosen for itself. If it is responded that money is nonetheless a good thing, because of what it can procure, Aristotle may readily agree; but then he will turn our attention to the things it purchases to determine whether or not they might constitute the final good.

Aristotle has similar reservations about the life of honour, and others as well (*EN* 1095b23–1096a4). Certainly living honourably is a good thing. Still, if we

LIVING WELL

seek honour as an end in itself, then we cede our happiness to the whims of others: people can be fickle and foolish, sometimes honouring the unworthy even while failing to honour the worthy. People can and do withdraw honour for any number of reasons, ranging from spite and jealousy to simple inattention. The final good, by contrast, is something 'genuinely our own and hard to take from us' (*EN* 1095b24–26). It appears, then, that honour is neither complete (*teleion*) nor self-sufficient (*autarkês*). Consequently, its presence, which may be specious, does not suffice to make a life lacking in nothing.

Perhaps the most challenging competitor for the status of qualifying as the final good is pleasure. After all, pleasure is a good thing, and it is chosen for its own sake and not for the sake of anything beyond itself. What is more, it is widely regarded as the best thing in life, as that which we in fact seek above all else. To appreciate Aristotle's attitude towards pleasure it is necessary and instructive to recognize the degree to which his ethical objectivism draws upon his underlying psychological theory. We have seen that Aristotle recognizes all living beings as ensouled, but also supposes that there is a hierarchy among the living, beginning with plants, which have only nutrition, through non-human animals, who add perception, to human beings, who are also rational. This explains why he uses harsh-sounding language regarding hedonists:

> The many, who are the most vulgar, seem to conceive of the good and happiness as pleasure, and accordingly love the life of gratification … In this way, they appear completely slavish, since the life they choose is the life belonging to grazing animals. But they do have an argument in their defence, since many who are powerful … are under the same impression.
>
> (*EN* 1095b16–23)

The hedonists regard themselves as cows, ruminating in their fields of pleasure, living for transient satisfaction and no more.

In rejecting the view of the many, Aristotle is not merely disparaging them with haughty rhetoric. He means, more literally, that those who seek only pleasure ignore that they are rational beings, and instead treat themselves as receiving only the sort of gratification possible for the unminded. In so speaking, Aristotle seems to be emphasizing physical over intellectual pleasure, and to be suggesting that pleasure-seekers situate themselves lower on the hierarchy of souls, because they are limiting themselves to sensuous gratification in the absence of intellectual activity.

LIVING WELL

Excerpted from *Aristotle*

One way to judge the correctness of Aristotle's case might be to envisage the possibility (perhaps not too far off) of a pink pleasure pill. You are offered the possibility of swallowing a pink pleasure pill. If you do, you are told, you will feel physical pleasure for the rest of your days. You will, however, do nothing, form no plans, pursue no goals, form no friendships, value no family ties. You will simply sit on a couch for the rest of your days, feeling pleasure, being fed, and being hosed off now and again. All of your days will be days of pleasure, though you will otherwise check out of all activity and all authentic interpersonal association.

Will you take the pink pleasure pill?

This question is, of course, not an argument but a simple appeal to intuition. Still, if you will refrain from taking the pink pleasure pill, that indicates that you are unwilling to regard at least this form of pleasure as the best life has to offer you. You think that your life holds higher possibilities, that the final good for human beings takes us beyond the realm of physical pleasure. Pleasure, again, is indeed a good. That is not in question. What is in question is whether it is the ultimate good for human beings. Aristotle's psychological theory provides reasons for adopting an ethical theory which does not enshrine pleasure in this position.

Thus far, then, we have seen Aristotle: argue that there is a final good for human beings; lay down criteria by which any pretenders to this role may be assessed; allow that we may regard the final good as happiness, or *eudaimonia*, only to insist that some conceptions of happiness, considered as the final human good, are superior to others; urge that subjective accounts of happiness be rejected in favour of objective accounts; and argue that given these constraints, three widely accepted accounts of happiness – the lives of moneymaking, honour, and physical pleasure – do not measure up.

His rejection of physical pleasure was especially significant insofar as it made free use of the metaphysics of human psychology developed in the hylomorphic framework of his *De Anima*. Aristotle at this point assumes himself justified in appealing to the essential features of human beings in his bid to explicate the best form of life available to us. He does not try to show that we should in fact desire the best form of life available to us, because he takes it for granted that people want what is in fact good for them and not what merely seems to be good without actually being so. What is really good for human beings, however, is determined by what human beings really are by nature. The nature of human beings is revealed, however, by reflecting on the teleological structures in terms of which the human function may be specified and understood.

LIVING WELL

Excerpted from *Aristotle*

8.3 Happiness and the human function

It may come as a surprise that human beings have a function. Computers and can-openers have functions, to compute and to open cans. We know that these sorts of artefacts have functions, and we have no trouble identifying what they are, for the simple reason that we gave them the functions they have. We designed them with specific purposes in mind. Matters are not so clear with the functions of human beings. To begin, Aristotle denies that human beings are designed by the activity of any form of intentional agent. Yet he does not suppose that this precludes our manifesting functions. Final causes occur in nature, he contends, even in the absence of intentional design. If that is so, then it should be possible to identify a human function, which in turn will provide a basis of a *functional account of human goodness*. That is, just as we may say easily that a good can-opener is a can-opener which opens cans well, we should likewise be able to say that a good human being is a human being who performs the human function well. The key, then, is to specify the human function.

Aristotle is aware that there may be doubts on this score, but thinks they can be met:

> But perhaps saying that the highest good is happiness will appear to be a platitude and what is wanted is a much clearer expression of what this is. Perhaps this would come about if the function (*ergon*) of a human being were identified. For just as the good, and doing well, for a flute player, a sculptor, and every sort of craftsman – and in general, for whatever has a function and a characteristic action – seems to depend upon function, so the same seems true for a human being, if indeed a human being has a function. Or do the carpenter and cobbler have their functions, while a human being has none and is rather naturally without a function (*argon*)? Or rather, just as there seems to be some particular function for the eye and the hand and in general for each of the parts of a human being, should one in the same way posit a particular function for the human being in addition to all these? Whatever might this be? For living is common even to plants, whereas something characteristic (*idion*) is wanted; so, one should set aside the life of nutrition and growth. Following that would be some sort of life of perception, yet this is also common, to the horse and the bull and to every animal. What remains, therefore, is a life of action belonging to the kind of soul that has reason.
>
> (*EN* 1097b22–1098a4)

LIVING WELL

Excerpted from *Aristotle*

Aristotle first notes that just as craftsmen have functions (a plumber plumbs, a programmer programs), so too do the parts of the human body; moreover, for things with functions, we judge goodness in functional terms (a good plumber plumbs well, a bad one poorly, and a good eye sees well, and a bad one poorly). Consequently, if human beings have a function, then we will know their goodness when we know their function. We know their function, contends Aristotle, when we know what is unique or characteristic (*idion*) about them – where, however, what qualifies as peculiar or unique will receive a technical treatment.

Aristotle's identification of the human good progresses in the form of his *function argument* (FA):

1. The function of any given kind F is determined by isolating the unique and characteristic activity of Fs.
2. The unique and characteristic activity of human beings is reasoning.
3. Hence, the function of human beings is (or centrally involves) reasoning.
4. Exercising a function is an activity (where, in living beings, this will be the actualization of some capacity of the soul).
5. Hence, exercising the human function is an activity of the soul in accordance with reason.

The function argument has proven controversial. Many regard it as wholly unpersuasive. Even so, some – though not all – of the difficulties critics locate in the argument result from misunderstandings.

Turning first to objections rooted in misunderstandings, it should be appreciated first of all that **FA** is not by itself attempting to prove that human beings have a function. On the contrary, Aristotle is at this point of his *Nicomachean Ethics* making free use of the hylomorphic analysis of human beings as substances articulated and defended in his *Physics, Metaphysics*, and *De Anima*. Central to this conception, as we have seen, is that kinds, including organisms, are individuated functionally in virtue of their having final causes. The argument presupposes, and does not attempt to shore up, Aristotle's four-causal explanatory schema; in so doing, it presupposes the framework of teleological explanation without trying to defend it anew. The function argument merely sets out to *identify* the function that Aristotle's teleology licenses him to assume that humans have.

One might acknowledge so much, or at least grant the teleological presuppositions of the argument, only to find it objectionable in its own terms. In particular, one may find **FA**-1 misguided. One may object, for instance, that merely identifying the unique or characteristic activity of a Fs hardly suffices to determine

LIVING WELL

the function of Fs. Obviously, the members of a given kind of entity might do any number of things uniquely without that activity qualifying as the function of that kind. Humans alone, it seems, drive around in big Cadillacs. Are we then constrained to conclude that driving around in Cadillacs constitutes the human function? Or again, perhaps only members of the human species sell sexual gratification for cash. If **FA**-1 entails that the function of human beings is prostitution, then **FA** is derailed even as it starts.

FA-1 has no such implication. In recommending that we seek the unique or characteristic activity of a kind, Aristotle intends something much stronger. First, that *some* members of a species engage in activities performed by members of no other species hardly makes that activity *characteristic* of the first species. Indeed, the single word rendered periphrastically as 'unique or characteristic' in **FA**-1, namely *idion*, is something we have already encountered in its technical role in Aristotle's theory of essence. In that connection, recall, an *idion* is a special sort of property, a necessary but non-essential property which flows from the essence of a thing, as for example it is an *idion* of human beings that they are capable of grammar, or capable of laughter, both traits explicable by the essence of human beings, namely rationality. In the present context, it is doubtful that Aristotle is appealing to the fully technical sense of this term, but it is clear that he means considerably more than what something happens to do, as a matter of contingent fact, uniquely. Rather, he means that we identify the function when we fasten on what it does characteristically, in a central way. Can-openers may also be paperweights, but it is not *idion* for them to play this role. If by chance it happened that all and only redheads were professional flute players, then it would not be *idion* of flute players that they have red hair. Minimally, we expect what is characteristic of a functionally determined kind *F* to be connected to the function and essence of that kind. This is why Aristotle recommends that when we are interested in identifying the function of human beings we turn our attention to what is peculiar or characteristic of them. Doing so will provide a road to essence, and so a road to final causation.

FA-2 asserts that when we go down this road we are able to identify the unique and characteristic activity of human beings as *reasoning*. Judged from a certain remove, one may wonder whether Aristotle is not being unfair to the non-human animals, since as cognitive ethology has taught us, plenty of other species engage in all manner of means–end reasoning; at any rate, many non-human animals can manipulate simple symbols, and the like. Now, in some

LIVING WELL

ways, it turns out, Aristotle proves sympathetic to these sorts of suggestions, since he is sufficiently impressed with animal behaviour to regard their perceptual activities as cognitively rich (*GA* 733a1); in another way, however, he is unsympathetic, since he accepts it as obvious that only human beings engage in natural philosophy, higher mathematics, and metaphysical speculation. One need not accept that there is a sharp distinction between higher and lower cognitive activities to accept that there is a relevant distinction to be made. Moreover, given Aristotle's functional determination thesis, if it emerged that non-humans, whether animals or aliens, really were rational, then they would simply share in the functionally determined human good. The plasticity of his account of kind membership automatically combats parochialism.

In any event, if we accept **FA**-2, then we accept that the human good is reasoning. **FA**-3 draws upon that commitment, but is by design non-committal on the question of how the exercise of the human function is to be understood, narrowly or comprehensively. That is, as stated, this interim conclusion holds that the function of human beings is *identical with* or merely *centrally involves* reasoning. Taken narrowly, this would amount to the claim that the human good is exhausted by rational activity, that the human good consists in the kind of theoretical activity we engage in when we do mathematics or philosophy.

Taken comprehensively, the human good might be understood much more broadly. Taken this way, we might understand the expression of reason to consist in the living of a well-ordered life, so that, for example, a life in politics might be conducted rationally or irrationally, where the rational execution of a political life would qualify as an admirable expression of the human good. At this juncture, we need not decide how Aristotle might be thinking of rational activity, noting only that FA eventuates in the conclusion that the human function is an activity of the soul conducted in accordance with reason, that is the living of a life which is an expression of the essential nature of the human kind, namely rationality.

Thus, Aristotle concludes, the human good consists in leading a fully and characteristically human sort of life. This conclusion highlights three distinctive features of Aristotle's account of human happiness: (i) happiness is an activity; (ii) happiness is objectively rather than subjectively determined; and (iii) happiness is forever rather than fleeting.

First, then, human happiness is a kind of living and is thus an activity rather than a passive state or affective experience: happiness is a doing rather than a being. That is, the best form of life is active rather than passive. This is yet another

LIVING WELL

reason why a human being would not, upon rational reflection, ingest a pink pleasure pill when offered: feeling pleasure is an affective state, whereas the best form of human life involves the execution of plans and projects. If it seems to you that you would not be living the best life available to you by sitting on a sofa, catatonic but feeling fine, then this may reflect some acceptance of Aristotle's thought that the best life consists in activity rather than in being affected.

The second distinctive feature we have already met, but can now better appreciate: the conditions of happiness are objectively given. We do not choose our essences. If an existentialist seeks to reverse this order by gamely claiming that 'existence precedes and rules essence', Aristotle will simply demur: we arrive in the world as rational beings, capable of engaging in the characteristic activities of our kind. Given that we have not chosen our kind, we have not chosen our ends; and given that we do not choose our ends, we do not choose our highest good. Of course, Aristotle has nowhere suggested that we cannot choose how we wish to pursue our good. Thinking of the human good comprehensively, we see that there are myriad paths to the expression of our essence, in philosophy, in the arts, in politics, in engineering, and so forth. Still, the end towards which these activities are directed is not a matter of choice; so, to be well chosen they must in fact be suitably end-directed. There are many ways for a harp player to play the harp well, but blowing into a tuba does not number among them.

Finally, then, these first two distinctive characteristics combine to give rise to the third. Aristotle expects happiness to be, if not exactly forever, then to range at least over very long stretches of life, perhaps over a whole life. He cites with approval a famous dictum of Solon's: 'Look to the end' (*EN* 1100a10–11), by which he means that we cannot stably judge whether someone is happy before the end of their life. While it is possible to judge an affective state episodically, the expression of an essence seems necessarily extended in time. That is, we can say, without fear of contradiction, we experienced pleasure last evening at 10.15 p.m. while eating dessert – but we cannot, according to Aristotle, sensibly say that we were happy between lunch and dinner, but not during either. Happiness, as the active expression of an objectively given end, is not like that. We will not judge someone to be a great violinist on the basis of a few notes well played, even if we think those notes greatly played; the judgment that someone is a great violinist requires more. Nor will we say that someone is a vegetarian for the period between breakfast and lunch if they have eaten no meat just then, especially if they have also eaten sausages for breakfast even while planning on a hamburger

LIVING WELL

for lunch. Any such judgment can be made only on the basis of a stable pattern of activity over a suitably long period of time. So too with judgments about happiness. If this suggestion bristles, this is likely at least partly due to patterns of speaking about happiness which are at variance with the objective character of Aristotelian happiness. We do indeed say, for instance, 'I was feeling happy before you phoned this morning.' Again, however, there is no point in quibbling about our unreflective manner of speaking. Still, to capture how Aristotle is thinking about happiness as the best form of life available for human beings, we might nonetheless note that it would be odd to say, 'I was leading a life which was the active expression of my essence as a rational being before you phoned this morning'. Of course, one could imagine a scenario where someone might be induced to utter such a sentence, but not readily. Aristotelian happiness (*eudaimonia*), to underscore our first two features, is neither fleeting nor a feeling; still less is it, then, any sort of fleeting feeling.

Once he has identified the human function in these terms, it is a short step for Aristotle to characterize the human good in his canonical expression (*EN* 1098a161–17):

- Happiness (*eudaimonia*) = df an activity of the soul expressing reason in a virtuous manner.

The sudden appearance of an appeal to *virtue* may be unsettling. So far we have been talking about the human good and our drive towards happiness without any mention of virtuous conduct at all. In fact, Aristotle's appeal to virtue in this connection is not at all out of place. In speaking of 'virtue' in this connection, Aristotle is thinking in the first instance of virtue in the sense of *excellence*. That is, Aristotle's word for 'virtue', *aretê*, makes it natural for him to think of virtue not only in the narrow, moral sense, but also in a broader non-moral sense also present in the semantic field of the English word 'virtue', though not as prominently as it is in the Greek *aretê* ('It was one of her great virtues as a general practitioner that her diagnostic technique was quick and flawless'). Thus, his account of the human good is equivalently the claim that it consists in the most excellent expression of the rational features essential to the human soul. The best life for human beings is a life expressing, in the most excellent manner, those features which make us distinctively human. Since happiness, or *eudaimonia*, is this highest good, we should expect it to be desired for its own sake, for the sake of nothing else while other goods are desired for its sake, and such that its presence renders a life complete and lacking in nothing. For these are, after all, the conditions laid down for

LIVING WELL

happiness, and in terms of which other contenders were set aside.

8.4 The virtues of character

A happy life is a life excellently, or virtuously, lived. It follows, Aristotle suggests, that an account of happiness will require an account of virtue, or excellence (*areté*) (*EN* 1102a5–7). Since, however, happiness is an expression of the faculties of the soul, the forms of excellence to be investigated do not extend to those pertaining to the body. An excellent body might be one with a good cardiovascular system or an efficient digestive tract, but these sorts of excellences are held in common with the non-human animals, and do not form the unique or characteristic (*idion*) feature of human beings. Rather, the forms of excellence or virtue requiring consideration are those pertaining to the human soul, which is a rational soul. An account of happiness will thus lead naturally into an account of the virtues belonging to the rational soul (*EN* 1106a16–26).

Which virtues are the virtues belonging to a rational soul? After all, even a rational soul is not rational in all of its aspects. It is a commonplace that the human soul is not purely or exhaustively rational. It is natural and easy to distinguish between reason and passion, between reason and desire or appetite, or between, in a popular idiom, the head and the heart. These contrasts are not the same; and they are hardly precise. On the contrary, each begs for clarification and defence, especially when agents appeal to such distinctions while seeking to excuse their bad conduct. ('I'm sorry. I don't know what came over me. I was so angry. I just wasn't thinking.') It is likewise natural to suppose, as Plato had urged in the *Republic*, that different parts of the soul can conflict and give rise to different sorts of actions. Appetite bids that I drink this water, while reason pauses to wonder whether this water is safe. Others have assailed this popular and philosophical thought by insisting, along with Hume, that reason and the passions *cannot* conflict, since reason is motivationally inert, whereas the passions by nature compel. Evidence for this way of thinking is supposed to follow from the fact that we can reason correctly that an almost imperceptible change in the standard of living in the first world could wipe out poverty in the third without there being even the slightest inclination to move in this direction. Reason calculates but does not impel; the passions motivate but do not reflect upon their ends.

These varying attitudes towards human motivation betray the shifting sands of moral psychology. Aristotle accepts a moderate position, eschewing the extremes of Hume while acknowledging the popular view that some parts of the

LIVING WELL

soul are rational and others not. It is easy to see that some parts are non-rational, given the theory propounded in *De Anima*: the nutritive soul is neither rational nor amenable to reason. Digestion is not irrational but simply non-rational. Still, suggests Aristotle, we may rightly identify another non-rational part of the soul, the seat of appetite and desire, which may indeed conflict with reason though it might also respond to reason and be integrated into its practical plans in a well-ordered life. He offers as evidence for this view that we speak freely of people who control their impulses and desires and contrast them with those who habitually succumb to the proddings of desire, only to experience regret and remorse after the fact (*EN* 1102a28–1103a3). He implies that unless we are prepared to be wildly revisionary about how we regard human motivational psychology, we should accept both rational and non-rational parts of the soul, and allow that these can come into conflict, but agree as well that they can equally be harmonized with one another in a unified agent.

These distinctions within the soul find correlates in our account of virtue. Given that we have identified one part of the soul which is purely rational and another which is non-rational but amenable to reason, we should anticipate that the kinds of virtue accorded to each will differ. In general, we have seen in *De Anima* that reason may be theoretical or practical (DA 431a8–17, 432b27–433a1, 433a14–16), and this Aristotle reaffirms in his *Nicomachean Ethics* (*EN* 1139a26–35). The theoretical sphere does not deal with action, but with understanding; the practical sphere, by contrast, concerns what is to be done, which action is to be undertaken and when. Hence, Aristotle concludes: 'Virtue is of two sorts, intellectual and moral' (*EN* 1103a14–16). Moral virtues are those virtues which pertain to character, but they are not confined to the non-rational part of the soul taken in exclusion from the rational. On the contrary, a person of virtuous character will subordinate the ends of her non-rational soul to those of her rational soul.

Focusing first on the virtues of character, Aristotle develops a general analysis of moral virtue with an eye on theoretical analysis not as an end in itself, but rather in order to determine the best route to becoming a good person. After all, he contends, the purpose of ethical theory is to help us become good (*EN* 1103b26–34). With this end in view, he proceeds by appealing to a distinctive doctrine rooted in the thought that virtue aims at a kind of habituation, in the inculcation of strong and deeply seated states of character, in a manner similar to what we find in craft production. If, that is, our goal is to produce good and decent

LIVING WELL

people, and their goodness and decency of character consist in their expressing stable virtues of character, we might look to the productions of crafts to see how best we might succeed. Aristotle observes that when we view a successful production of some craft, say a beautiful table masterfully executed by a journeyman carpenter, we find ourselves agreeing that a kind of equilibrium or balance has been reached: adding or subtracting anything at all would only detract from the product *(EN* 1106b8–16). So too, perhaps on the basis of this admittedly slender analogy, Aristotle argues that when realized, virtue achieves a mean between excess and deficiency.

Tying together some of these strands, Aristotle offers a general account of moral virtue, or the virtue of character:

> Virtue is a state of the sort which issues decisions, consisting in the mean relative to us, determined by reasoning of the right sort, that is the reason in terms of which a wise person (the phronimos) would determine it. It is a mean between two vices, one of excess and one of deficiency.
>
> (*EN* 1106b36–1107a6; cf. 1138b18–20)

Although he does not advance a tidy argument on behalf of this account, Aristotle does offer considerations on behalf of each of its components. In an effort to understand this account, we must consider at least briefly each of these:

- The first component is that virtue is a state (*hexis*). Aristotle argues that virtue must be either a feeling (*pathos*), a capacity (*dunamis*), or a state (*hexis*). It cannot be a feeling, since people are regarded as excellent or rotten on the basis of their manifesting virtue or vice, but not insofar as they have feelings of one sort or another. Further, virtue cannot be a mere capacity, since we are endowed by nature with capacities and become virtuous only by exercise and habituation. Hence, virtue must be a state (*hexis*), an acquired but entrenched condition of character, achieved through guided development and habituation (*EN* 1105b20–1106a13). Here too the appeal to crafts is apposite. No-one learns to be a journeyman carpenter in a day; and once this state is achieved, we can expect the carpenter to remain a journeyman for a long while, if not forever. So too with virtue: it is a stable state (*hexis*) not readily lost once acquired.
- Virtue is the sort of state which issues in decisions (*hexis prohairetikê*). Aristotle is speaking fairly technically here, employing a term which he elsewhere indicates involves our being in a state which presupposes prior

LIVING WELL

deliberation (*EN* 1112a14–16). He is not suggesting, however, that virtuous conduct requires deliberation immediately preceding each and every action. On the contrary, virtuous action flows directly from an entrenched state. He means that a virtuous state is one which, having been guided by deliberation in its inculcation, is the sort which eventuates in decisive action. Again, the journeyman carpenter does not pause to deliberate about how best to saw when building. That sort of activity flows directly from the settled expertise already developed.

- Virtue is determined by reasoning of the right sort (*orthos logos*), reasoning that can eventuate in a general direction for conduct in a general situation, though not in a fine-grained or determinate rule for all situations (cf. *EN* 1138b18–1140b24).
- This sort of right reasoning is precisely what a wise person, or person of understanding and practical wisdom (the *phronimos*), would arrive at in the situation in view. Such a person is able to grasp what is in fact true about what is good or bad for a human being, and so will not be liable to imprudent confusions on this score. Aristotle is not saying that the *phronimos determines* what is right by fiat, but that since the wise person characteristically *recognizes* what is right, it is sensible to follow his lead (cf. *EN* 1140a25–b6).
- Finally, virtue is a *mean* (*mesotês*) between extremes, but only relative to us. In speaking of a mean *relative to us*, Aristotle suggests that an agent must look to herself and her context in making a determination. Thus, it would be wrong to rely upon a purely quantitative formula. If six is the mean between ten and two, we should not infer that we should eat six pieces of pizza, on the grounds that ten would be too many and two too few. What we should eat will depend upon facts peculiar to us, how much we weigh, how quickly we metabolize and so forth. Perhaps Milo the wrestler should eat six pieces of pizza, because that would be healthy for him; this would not be the mean amount for most people. Carried over to the virtues of character, Aristotle's suggestion is that there is not, for example, an unwavering amount of righteous indignation suitable for all agents in all contexts. High indignation is excessive when a waiter gives us the wrong spoon for our grapefruit, but not so out of place if the same waiter without provocation tells us that he might enjoy molesting our daughter.

Among these components, the most distinctive is the doctrine of the mean, which

LIVING WELL

accordingly requires further development.

Aristotle notes that we do not in every instance seek the mean, or seek the mean under every possible action description. Some actions are base, come what may: we do not practise adultery with the right neighbour at the right time and in the right amount. As its very name suggests, adultery is a vice (EN 1107a9–25). Aristotle's point here has both substantive and non-substantive dimensions. On the non-substantive side, he is merely pointing out that a mean exists only relative to some descriptions of actions and not others. Still, a judgment about which descriptions are to be employed already reflects a judgment about what is to be regarded as good or as otherwise – as adultery is always regarded as base. Perhaps, though, we join him in this presupposition when we argue, for example, whether a killing was a case of manslaughter or a murder, the implication being that if it was a murder the killing was more reprehensible then if it were mere manslaughter. Even so, there remains some difficulty about when it is appropriate to select descriptors which presuppose that an action is so thoroughly vicious that it is nowhere on a continuum upon which virtue sits as a mean.

That acknowledged, Aristotle's doctrine of the mean is best understood in relation to the individual virtues, as he himself insists. In presenting his theory, Aristotle finds it necessary to make recourse to neologisms and appropriations from ordinary language. This he regards as unproblematic, since he notes that in some cases, the extremes are nameless (*EN* 1107b2). This may simply derive from the fact that we rarely or never encounter people deficient along some dimensions. In any event, he will suggest that where rashness and cowardice are the deficiency and the excess, courage is the mean; between self-indulgence and self-deprivation, moderation is the mean; where great sums of money are controlled, between the excesses ostentation and niggardliness is the mean of magnificence; but where smaller sums are concerned, between wastefulness and stinginess lies generosity (*EN* 1107a32–1108a31). In these and like cases, Aristotle thinks it is in principle possible to place virtuous action along a continuum, even if the ends of the continuum are not recognized in popular discourse.

Consequently, Aristotle's doctrine of the mean has come in for criticism, sometimes on the grounds that it is forced or artificial, or perhaps insufficiently general. The first sort of criticism in this direction carries little weight. The bare fact that we lack names for some excesses or deficiencies matters little unless it can be shown that the only excesses or deficiencies of character are those that we have happened to notice and name. A second sort of criticism is more

LIVING WELL

consequential. To see the problem, consider the virtue of honesty. We would normally think of honesty as a virtue yet there seems to be no obvious excess to which the corresponding deficit is a vice. That is, the vice associated with honesty is lying, its opposite, not some defect on a scale on which honesty is the mean. Thus, even if one were to manufacture an excess, say painful truth-telling ('My, you've gained weight'), there seems to be no non-forced continuum along which lying is at the other end. Now, how serious a problem this might be depends in part upon the range of virtues we are prepared to entertain. The sort of honesty Aristotle discusses in the *Nicomachean Ethics* is only a kind of self-regarding honesty, which seems reasonably well suited to his preferred treatment. This is honesty restricted to one's own accomplishments, where the excesses are boastfulness and self-deprecation (EN 1108a20–23). It is, however, a difficult matter to determine which virtues we should in the end be prepared to entertain. In one direction, it would be inappropriate to allow Aristotle to select only those amenable to treatment in terms of his doctrine of the mean; in the other, without external warrant, we would be premature in concluding that his framework topples because it cannot handle some seeming virtues to which it is ill-suited. In either case, however, the onus is upon Aristotle to supply a legitimate decision procedure to deal with disputed cases. Otherwise, at the very least he will be guilty of an unacceptably blinkered parochialism.

8.5 A puzzle about *akrasia*

The virtues of character do not exhaust human virtue, since there are equally virtues of intellect, belonging to the rational part of the soul, which are still to be considered. Aristotle devotes *Nicomachean Ethics vi* to this task only to give way in the following book to a tangled and engaging discussion of *akrasia – incontinence or weakness of will*, or perhaps simply *lack of self-mastery*. Recall that Aristotle had insisted when setting up his discussion of the virtues of character that there are two parts of the soul, one rational and one amenable to reason. One bit of evidence for that distinction appealed to common experience, that we sometimes find ourselves at variance with our own reasoned decisions, to the point where we find ourselves doing things we had determined not to do (*EN* 1102a28–1103a3). Common as they may be, such experiences are puzzling, in part because they raise questions about the relations between the different parts of our souls, or our selves. Suppose I determine to exercise today. I don my exercise clothes and head

LIVING WELL

towards the gymnasium – but then decide along the way to step into a pub for a quick drink before working out. I end up socializing rather than exercising, and tomorrow morning regret that I failed, again, to do what I had determined to do. One might well wonder: if I earlier determined to Φ and now regret not Φ-ing, then how am I related to the person who decided to Ψ instead of Φ between then and now? Surely I am the same person (hence my regret), and in between I simply chose of my own unforced will to abandon my earlier determination to Φ (but, then, why my regret?). Perhaps I wish I were not the sort of person I am; but then again it is all along open to me not to be such a person.

These sorts of questions are salient for Aristotle both because he hopes to capture the phenomena of our lives as we lead them and because he needs to show how the rational and non-rational faculties of our souls intersect to make fully flourishing human life practicable. Aristotle has some difficulties characterizing *akrasia*, and displays a bit of ambivalence about how best it should be conceived.

On the one hand, he is critical of Socrates, who had argued in the *Protagoras* against the possibility of *akrasia*. Socrates had argued that at least against the backdrop of a certain kind of highly unified agency *akrasia* is impossible because unintelligible. If, for instance, I am a hedonist who is forever interested in maximizing my own pleasure, then it makes no sense for me first to determine that Φ is the pleasure-maximizing activity, but then to do Ψ instead of Φ, on the grounds that Ψ has overwhelmed me with its promise of pleasure. Still less is my *ex post facto* justification easily intelligible. I was weak, I say, and succumbed to the lures of some beckoning pleasure. To the extent that these explanations are unintelligible, *akrasia* must be impossible.We are, then, mistaken in our own self-characterizations when we claim to be akratic.

Socrates' argument in this direction, Aristotle cautions, 'contradicts the appearances' (*phainomena*)(EN 1145b27–28). That may seem fair enough: surely we are sometimes weak-willed. Indeed, for many of us, the akratic cycle is regrettably familiar: we resolve to pursue a course of action *a* in preference to *b*, because we believe that *a* is, all things considered, preferable to *b*, and yet then, at the moment of action, choose *b*, only later to engage in harsh self-recrimination and remorse, followed by renewed resolve to be stronger and better at the next opportunity. Surely, if he had wanted to deny the existence of this sort of experience, then Socrates would have had to explain away a fair bit of our lived lives. This is what Aristotle intends when criticizing him for contradicting the way things appear. Yet,

LIVING WELL

arrestingly, at the end of his treatment of *akrasia*, Aristotle comes around to a much softer judgment of Socrates. In a certain way, he says, his own account seems to give Socrates what he was looking for, which was that knowledge cannot be dragged about like a slave by mastering passions (*EN* 1145b324–25, 1147b15). One approach to understanding Aristotle's position is, then, to determine how Socrates is and is not right about *akrasia*.

The matter is complicated because we have two layers of interpretation interacting, namely our view of what Socrates held in the *Protagoras* and Aristotle's presentation of him, perhaps drawing upon that same dialogue. As Aristotle represents him, Socrates denied the phenomenon of *akrasia* by treating all cases of weakness as involving cognitive failure. We will mainly defer to Aristotle's presentation of him, since in the present context we are trying to work out Aristotle's view of the matter. According to Socrates, we should not assign the causes of our weakness to a failure of will, or to an overpowering desire of any kind, but to a miscalculation. Indeed, relative to a certain group of background assumptions, this may seem just right. Suppose that we are highly unified, in the sense that we submit all decisions to a single, seamless faculty governing action. Further, if we are egocentric hedonists always concerned with our own pleasure maximization, and always focusing our attention on the single sort of pleasure there is, then it is hard to see how we might go wrong – unless we somehow fail to understand the likely results of our actions. It is as if we were dedicated stock market investors, who, having determined how best to maximize profits, nonetheless decided to invest our money in stocks we expected to be substandard performers. Such conduct would be odd. Minimally it would require some explanation. After all, we would have no motive to engage in such conduct in the circumstances envisaged. More likely is the thought that if we in fact purchased the poorly performing stocks, the only plausible explanation for our doing so resides in some miscalculation that we made along the way. That, though, is a cognitive error, and not any form of weakness of will.

It is fairly easy to see how Aristotle thinks this Socratic picture has gone wrong. It will prove less easy to see how it has gone right. To begin, according to Aristotle, the Socratic picture relies upon a false moral psychology, according to which we are in fact highly unified agents.We are not. As we have seen, our souls have rational and non-rational facets and these can come into conflict in compelling us to act. So, in the first instance, Aristotle objects to the underlying psychology giving rise to the putative impossibility of *akrasia*.

LIVING WELL

Aristotle argues that there are further complexities which must equally inform our account. To begin, two background distinctions must be observed. First, we can speak of both *having* and *using* knowledge, a distinction already familiar from De Anima (*DA* 417a21–b1), where it was given in terms of first and second actuality (*EN* 1146b31). Claire might have the knowledge that there is a detour along her customary route home, but not be using that knowledge just now, in the sense that she is not at present thinking about it, for any number of reasons. Those reasons are inconsequential if she is not now driving home, because she is at work, as a doctor attending to her patients and concentrating on their care and treatment. They become consequential if, when driving home, she is so distracted by wondering whether she has ordered the correct treatment for a patient that she does not make use of her knowledge; she will likely regard herself as blameworthy while sitting in traffic, lamenting that she *knew* there was a detour to be avoided.

The second preliminary distinction is a bit more complex, involving what appears to be in Aristotle's mind a rational reconstruction of the antecedents of our action. Each time we do something intentionally, he suggests, we may regard our action as preceded, at least implicitly, by a kind of *practical syllogism*, made up of a universal and a particular premise. The universal premise commends such and such a goal to be pursued, for example that sweet things are to be eaten (*EN* 1144a31–33). The particular premise locates the actor in a situation wherein the universal premise applies, for example in a situation where one perceives that this piece of cake is something sweet. This seems at best a rational reconstruction because we do not actually rehearse such a syllogism for ourselves each time we eat a piece of cake. Even so, it is plausible in a broad range of cases that some such reconstruction is both possible and apt.

Armed with these two distinctions, the basic outline of Aristotle's approach to *akrasia* is clear, though thereafter matters become hotly disputed. His basic thought is just that, combining these two preliminary distinctions, knowledge failures may take several different forms. One might: (i) have but not use knowledge of the universal premise; (ii) have but not use knowledge of the particular premise; or in principle (iii) have and use knowledge of both premises, but fail to use them concurrently. This last suggestion may be a bit surprising, but the idea has a natural logical analogue. Raphael may know that all mammals have lungs and that this dolphin is a mammal, yet find himself surprised that the veterinarian proposes to perform a lung transplant operation on this dolphin, because he somehow failed to connect the two bits of knowledge he was using,

LIVING WELL

and somehow did not appreciate that this dolphin has lungs. If that is possible, then his knowledge failure is somehow a gestalt matter rather than a local one. That is, it involves the interaction of discrete items of knowledge rather than any one item of knowledge taken in isolation.

Aristotle draws attention to this sort of gestalt affair (*EN* 1147a31–b5), and also to both of the premises of a practical syllogism taken individually, sometimes emphasizing knowledge failures pertaining to the particular and other times pertaining to the universal. Wherever one should locate the knowledge failure involved, *akrasia* is possible, he suggests, because of one's 'knowing and not knowing' (*EN* 1147b17–18). To this extent, Socrates is vindicated after all: *akrasia* does involve a kind of knowledge failure, if not the simple sort of knowledge failure he had envisaged (*EN* 1147a14–19).

Beyond so much, Aristotle's treatment of *akrasia* resists easy interpretation; it is also consequently difficult to assess its defensibility. There is little scholarly consensus regarding the precise contours of his view, though this may be due in part to the unclarity both he and we have about the phenomena under consideration. It is not special pleading on Aristotle's behalf to note that puzzles about *akrasia* admit of a range of formulations, some of them arcane and removed from experience and others striking in their simplicity. Hence, if some of the difficulty with Aristotle's treatment results from his own hesitance and unclarity, it seems fair to conclude that some also results from the permanently puzzling character of the phenomenon.

8.6 Friendship

Aristotle discusses many virtues, both moral and intellectual, throughout his *Nicomachean Ethics*. One sort of virtue, or concomitant of virtue (*EN* 1155a3–5), merits special treatment because of its tendency to correct a misimpression we might form about Aristotle's ethical theory. The misimpression is that Aristotle's theory is thoroughly egoistic: we have been focusing on happiness (*eudaimonia*) and the best way to secure it. It might be natural to conclude on this basis that Aristotle's ethical theory begins and ends in an account of self-regarding attitudes. The corrective to this misapprehension is Aristotle's treatment of friendship (*philia*).

One might well ask, in a narrowly self-interested vein, why we should want to have friends, if having friends requires us to care about their well-being, even when our doing so in turn requires us to sacrifice our own interest to theirs. According to one form of egoism, even if we think of them as necessary for our own

LIVING WELL

happiness, it might nonetheless seem that friends are best regarded as mere instruments to our own pleasure, toys to be played with when they suit our interests but shelved when they do not. One question, then, concerns whether Aristotle's ethics permits a broader, less instrumental attitude towards friendship.

Aristotle identifies different kinds of friendship (*EN* 1156a6–b33), some of which might seem to expect nothing more than this sort of unseemly instrumentalism:

- Friendships based on *utility*, where a bond is formed primarily on the basis of mutual benefit, of the sort characteristic in ongoing business relations.
- Friendships based on *pleasure*, where the basis of the relation is shared pleasures, as when witty people delight in exchanging clever remarks.
- Friendships based on *goodness*, complete or perfect friendships, where two people equal in virtue care for one another, each for the sake of the other and form their friendships on the basis of character.

Aristotle observes that the first two forms of friendship, which he regards as secondary, are easily dissolved and tend to disappear when the source of the friendship dries up.

If these sorts of relationships are instances of friendship, then friendship does not require concern for another for her own sake (cf. *EN* 1155b31–1156a5, where Aristotle nonetheless seems to imply that all friendship requires such regard). For neither friendships of utility nor friendships of pleasure seem to require any other-regarding consideration on the parts of those who enter into them. It is sometimes suggested that the oddness we may feel in this results from the fairly wide compass Aristotle has in view for what he calls *philia*, which extends beyond friendship in a familiar modern, social sense. The translation, however, does not seem inapt, since we equally speak of friendships in business relations ('I have a friend in shipping who can check the status of the order') and pleasure-based relations ('Marcus was the sort of friend I called when I was feeling blue and wanted to forget my troubles'). The main concern with such friendships is that they are secondary forms of friendships, as Aristotle suggests they are: though they are useful, or even necessary, for commodious living, they do not represent Aristotle's primary interest in friendship, which he restricts to the finest kind, friendships based on goodness. In its finest form, friendship endures as long as virtue endures; but since virtue is a stable state of character and is essentially extended in time, true friendships are not easily dissolved.

Indeed, in perfect friendship, we expect friends to regard one another as

LIVING WELL

second selves. Partly on this basis, Aristotle argues that we have reason to love others as we love ourselves – and we do have reason to love ourselves. Once we distinguish appropriate self-love, founded in a correct view of the self as a rational being, not as a self-involved seeker of pleasure, money, or honour, we have reason to regard ourselves as the bearers of intrinsic worth (*EN* 1168b11–1169a7). In perfect friendships between equally virtuous partners, however, one friend will share the other's character, so that what one cherishes in oneself one will also recognize in another. The good loved in oneself will then be equally realized and loved in one's second self. There being no relevant distinction between these forms of goodness, one friend, suggests Aristotle, will have cause to sacrifice goods, wealth, even life, for another. This, of course, is the crucial cross-over, or attempted cross-over, from self to other, and so from narrow egoism to an undeniable form of altruism. Perfect friendship implicates one friend in other-regarding conduct towards another. Aristotle's ethics thus avoids the egoism with which it is sometimes charged; true friendship is at once self- and other-regarding.

Aristotle buttresses this suggestion by reaching back to the original conditions on the highest good, arguing that friendship is necessary for self-sufficiency, that condition which when satisfied yields a life lacking in nothing (*EN* 1097b6–16). We will, then, be motivated as eudaimonists to seek our own happiness; we achieve human flourishing, however, only in the company of indispensable friends. When we have friends of great goodness and character, we recognize their worth antecedently: they are not good because they are our friends, but are our friends because they are good, and manifest those traits we rightly recognize as good in ourselves. To counter that we are thus using such friends for the sake of our own happiness confuses perfect friendship with friendships of utility.

In fact Aristotle's treatment of friendship's basis for altruism has two discernible strands, each perhaps relying on the other. He does not offer them as discrete arguments, though they do seem to draw upon importantly different considerations. They are best presented in tandem, so that the distinct wellsprings of each may be emphasized. This process also helps highlight an easily over-looked component of Aristotle's arguments, namely that they draw freely upon what he now regards as settled doctrines defended in his metaphysical and psychological theories.

The first argument takes seriously the language of friends as *second selves* (SS) (*EN* 1107b5–14):

LIVING WELL

1. If we are fine and virtuous, then we regard ourselves with proper self-love – proper because virtuous features are indeed worthy of love.
2. If those features are worthy of love as they occur in us, then they are no less worthy of love if they occur in our friends, who are our second selves.
3. Because they are our equals in virtue, our friends will indeed manifest the same fine features we ourselves manifest.
4. Hence, the fine features manifested by our friends are worthy of love.
5. If their features are worthy of love, then this gives us reason to care for our friends because of who they are.
6. Hence, we have reason to care for our friends because of who they are (*EN* 1156a19–11, 1156b10, 1156a17–18).

On this sort of basis, Aristotle concludes that 'just as each person's own being is choiceworthy, so too is a friend's being choiceworthy' (*EN* 1170b7–8).

The first premise (SS-1) reaffirms that proper self-love is perfectly virtuous. If we falsely deny our rational worth, then we are self-deprecating; if we exaggerate our worth, then we are self-aggrandizing braggarts. If it is indeed true that we are in a condition answering to the criteria set for the best human life, then trivially we have reached some condition which is good in itself, and appropriately acknowledged as such. SS-2 contends that virtuous traits are not enhanced by being our virtuous traits, nor diminished by being the virtuous traits of someone else. This is all the more emphatically so if my friend is my second self. Now, it is tempting to insist at this point that talk of 'second selves' is oxymoronic: a self is necessarily an individual, and there can be at most one of each. It is doubtful that Aristotle seeks to deny this. Rather, friends of equal virtue are tokens of a type; and the type is something worthy of love. It is difficult to determine why it should not be arbitrary to love one betokening over another. If a serious composer rightly regards his masterwork as realizing great beauty, but recognizes this same beauty in the composition of another, there seems little room for him to insist that the beauty of his work is more beautiful, or more valuable, because it is manifested in his work.

It is important to realize when assessing SS-2 and the premises which follow that in this argument Aristotle is relying upon his general account of human happiness, as objectively given and determined by the essence of human beings. If we recall Aristotle's objective account of happiness at this point, then we appreciate that judgments about happiness are judgments about human flourishing. If we think that human flourishing is a good thing, then we find it good

LIVING WELL

in our flourishing friends no less than in ourselves. It follows, of course, that when we show concern for our friends, we are not interested in helping them secure the ends of their desires, whatever those may be. On the contrary, if our friends desire things inimical to their flourishing, then we tell them so, precisely because they are our friends and we care for them.

The second strand in Aristotle's defence of friendship also reaches back to his general account of happiness and its metaphysical underpinnings. He argues that one condition set for the best life is especially significant when we come to ponder the point of friendship. The final good for human beings must be self-sufficient (*autarkês*), such that its presence suffices to make a life lacking in nothing (*EN* 1097b6–16). Aristotle now argues boldly that one cannot be self-sufficient without friendship:

> If being is choiceworthy in itself for the person who is blessed, because it is naturally good and pleasant, and if the being of his friend is closely similar to his own, then his friend too will be choiceworthy. Whatever is choiceworthy for him he must possess, since otherwise he will in this way be lacking in something. Hence it is necessary for anyone who is going to be happy to have excellent friends.
>
> (*EN* 1170b14–19)

In some ways, this argument draws upon Aristotle's view that friends are second selves; but it adds a stronger claim as well. The claim it adds is that one who lacks friends lacks selfsufficiency (LSS):

1. If S does not possess a choiceworthy friend, then *S* lacks something choiceworthy.
2. If S lacks something choiceworthy, then *S* is not self-sufficient
3. If S is not self-sufficient, then *S* is not happy.
4. Hence, if *S* does not possess a choiceworthy friend, S is not happy.

LSS-1 seems to draw upon Aristotle's reflections on the interaction between proper self-love and the recognition of the grounds of that love as manifested in another. He once again reverts to the general framework of his objective conception of happiness by recalling that the best life, whatever it may be, will be one which is selfsufficient. If friends are necessary for self-sufficiency, then friends

are equally necessary for happiness.

In one way, LLS-4 may not seem to take us from narrow egoism to some

LIVING WELL

form of altruism. After all, once someone has a friend, it may be observed, she may be happy; should the loss of that friend threaten unless sacrifices are made, then that friend will only need to be replaced by another. So, the demand for self-sufficiency, even thus interpreted, remains compatible with an unseemly instrumentalism.

Aristotle's attitude towards this sort of criticism is multitiered. To begin, he implies that this sort of complaint may simply betray an especially fatuous kind of psychological egoism: it seems to presuppose that it is always possible, or perhaps even necessary, to regard others in wholly instrumental terms. Aristotle doubts this, since once it is agreed that a friend, because virtuous, has attained some objective intrinsic value, it becomes difficult to fathom why this should be set aside when we move to act, or indeed how it could be set aside – if, that is, we have formed a perfect friendship with that person. If we have reason to be virtuous, and friendship is a virtue, then we have reason to develop perfect friendships. Having developed such bonds, we will act for the sake of others as an expression of our friendship towards them. If we are thinking of their usefulness to us, then we are also thinking of them not in terms of perfect friendship, but in terms of friendships of utility. It seems unnecessary to agree that all friendships must be restricted to mere friendships of utility. Moreover, it seems implausible that a human being flourishing fully in Aristotelian terms would be disposed to regard all others – all intimate friends, all family members, all whom we love – in such narrowly instrumental terms.

Part of the reason that this seems implausible to Aristotle is that we are likely to have appreciably different sorts of affective responses to friendships based on utility and friendships based on goodness. In order to illustrate the sorts of affective responses we can expect perfect friends to evoke from one another, Aristotle frequently appeals to the sort of spontaneous tender regard a mother has for her children (*EN* 1159a28, 1161b27, 1166a5–9). It is a commonplace that parents willingly suffer and sacrifice for their children's well-being. From the detractor's point of view, perhaps the behaviour of parents is irrational. From Aristotle's, it represents the normal human affective response to an object of love.

8.7 The final good for human beings reconsidered

After completing his accounts of the virtues, Aristotle returns in he last book of the *Nicomachean Ethics* to review the best life for human beings, as he had in its first book. Although it begins as a familiar summary, the recapitulation carries a

LIVING WELL

remarkable surprise. As Aristotle recounts his view, he introduces elements not only left unmentioned in his earlier treatment but so singular and distinctive that they threaten to contradict the earlier account directly. To some scholars, the contradiction is so plain and palpable that it shows clearly that the tenth book of the *Nicomachean Ethics* cannot form part of a single work with the preceding nine. To others, matters have seemed less dire; although there does seem to be some tension, it is possible to reconcile what is said in these different parts of the work simply by attending to Aristotle's presentation of the issues.

The problem arises most directly when Aristotle revisits his conception of the best life:

> If happiness is an activity in accordance with virtue, it is reasonable that this will be the supreme virtue; but this will be the virtue of what is best. Whether this is reason or something else which seems by nature to rule and to lead and to have thoughts of things fine and divine – be it itself divine or the most divine element within us – its proper activity will be complete happiness. As has been said, this activity is the activity of contemplation. This would agree with what has been said before, and also with the truth.
>
> (*EN* 1177a12–19)

It is surprising to find Aristotle contending that the view expressed here coheres with what has been said elsewhere. For, on the contrary, whether or not what he says here agrees with the truth, it does not seem to agree with what has been said before, because it was not said before that the human good consists in contemplation. Rather, having divided the rational soul into the rational and non-rational, the bulk of the *Nicomachean Ethics* has pursued discussion of the moral virtues, or virtues of character, followed by a comparatively brief discussion of the theoretical virtues. If the expression of virtues of character is now excluded from the realm of happiness, then the current claim not only fails to cohere with what has been said earlier, but cannot even be made to reconcile with it.

Put into sharper relief, the problem Aristotle faces may be seen as his accepting the following inconsistent pair of propositions, one an encompassing conception of the good and the other a narrow conception:

- *An Encompassing Good*: The human good consists in the expression of human virtue, where human virtue includes a broad range of activities, encompassing the full range of moral and intellectual virtues.

LIVING WELL

- An Exclusive Good: The human good consists in the expression of human virtue, where human virtue is limited to the finest intellectual virtue, namely contemplation.

Put in these terms, if Aristotle maintains that the human good is an encompassing good throughout the bulk of the *Nicomachean Ethics* only to conclude by asserting the exclusive conception in its last book, then he is in an uncomfortable situation.

Even before wondering about matters of consistency, however, the exclusive conception of the human good causes concern in its own terms. Surely, one may fear, the exclusive conception threatens to be excessively narrow. After all, the virtuous person is expected to have friends, and is expected to be just, and to do so because her human fulfilment consists in the expression of virtues which are ineliminably social in character. Elsewhere Aristotle will assert, in keeping with this broad conception of the human good, that humans are *by nature* political animals, that indeed their essential traits lead them to form social associations (*Pol.* 1353a7–18, 178b15–30). If it were now to turn out that the human good consists solely in contemplation, and that all we do we do for the sake of contemplation, then nearly all of our actions will ultimately be directed at something solitary and fundamentally asocial, something more god-like than human.

In fact, Aristotle does seem to assert that we should strive to be as god-like as possible (*EN* 1177b26–1178a2), where he conceives god's activity as restricted to a remarkably austere form of self-referential contemplation (*Met.* 1074b29–30). If all is done for the sake of those rare moments when we can ourselves reach up and cross the intellectual threshold into the realm of the divine, then we are rarely flourishing, since our moments of contemplation will only seldom punctuate our otherwise animal lives of eating, drinking, and socializing. Moreover, it will seem on this narrow conception that nothing but this lofty form of activity will be intrinsically valuable, since outside of contemplation all will be done for the sake of something beyond itself. So much then seems to ignore that we are human animals, preferring instead to pretend that we are human godlets.

Now, the general tension encoded in these broad and narrow conceptions of the human good has spawned a vast literature.

Here we can only gesture towards two sorts of resolutions, the first giving way to the complexities of the second. These resolutions attempt to avoid concluding directly that Aristotle has contradicted himself. That is, of course, a possibility. Another possibility of the same general tendency would not ascribe a contradiction to Aristotle but allow that he must somewhere simply have changed

LIVING WELL

his mind, as many of us often do, since the views are inconsistent with one another, and we in any case have independent reason to think that the last book of the *Nicomachean Ethics* cannot form one part of a unified work whose other parts include the first nine. Whatever such independent reason may amount to, we should appreciate that the problem introduced here in principle admits of a number of resolutions.

This may be in part because the (seeming) contradiction between the encompassing and exclusive conceptions of human flourishing gives rise to a range of distinct problems. The first sort of resolution is rather deflationary, though it need be none the worse for that. It bears immediate notice that Aristotle is aware of some tension in this direction and is prepared simply to rank forms of happiness. After ending *Nicomachean Ethics x 7* by insisting that the life of contemplation 'will be the happiest' (*EN* 1178a8), he opens the next chapter by observing:

> Second happiest is the life led in accordance with the other sort of virtue; for activities of this sort are human. For we do just things and courageous things and the other kinds of things in accordance with this sort of virtue in relation to one another … and all of these appear to be human.
>
> (*EN* 1178a9–14)

One easy thought would then be this: happiness admits of degrees, the best happiness is contemplation, but the second best happiness, which is genuine happiness all the same, is the sort which embraces all forms of human virtue, intellectual and moral alike. There will then be a threshold to cross for happiness, above which some will be happier than others, though all will be, so to speak, fully happy. Suppose that in order to attain first-class honours a student must score above 95 per cent in her final examinations. One student scores an admirable 95.1 per cent and another an astonishing 99.9 per cent. Both have, fully and completely, earned first-class honours; neither is more first-class than the other. Still, there is a fair sense in which one has achieved more than the other, and is thus more honourable. In the case of human happiness, judgments of scale are fully appropriate, because happiness consists in actualizing a functionally specified final good, and functional kinds are scalar kinds.

Of course, this sort of deflationary resolution may be fine as far as it goes, but it does not go far enough. That is, even if correct, it fails to address an underlying concern regarding the question of whether actions done *for the sake* of

LIVING WELL

happiness must be regarded as having merely instrumental value. For surely actions done for the sake of an end beyond themselves may also be valued in themselves as well. Moreover, one might expect a human life to comprise all manner of good activities, things done for their own sake, and not all exclusively subordinated to one unified goal. In this sense, our worry about exclusive versus embracing conceptions of the good gives way to a worry which may have been nagging us already from the very first sentences of the *Nicomachean Ethics*: as the work opens, we learn that every action aims, ultimately, at some one good. What, then, is the relation of things done on behalf of this good and the good itself?

Already in the first book of the *Nicomachean Ethics* Aristotle had maintained that the human good is 'an activity of the soul in accordance with virtue (or excellence, *aretê*), and if there are many virtues, then in accordance with the best and most complete' (*EN* 1098a16–18). Looked at one way, this may be paraphrased as ' ... and if there are many virtues, then the human good will be an activity identified with the one which is best and most complete among them'. Looked at another way, this may mean ' ... and if there are many virtues, the human good is an activity identified with the best, most complete virtue'. According to the first paraphrase, there is to be some one virtue, selected as best from among them all, in whose expression the human good will consist. According to the second, the best virtue will not be thought of as competing with other virtues. Rather, if there is a plurality of virtues, the most complete package of them will be the human good. This is roughly the difference between saying that if there are many beautiful flowers, what is best will be the single most beautiful flower among them as *opposed to saying* that what is best will be the most beautiful bouquet of them all, which will surely feature the most beautiful among them.

Which does Aristotle intend? The matter is disputed, and in a certain way turns on a partly linguistic matter concerning the question of what it means to say that *S* does *a for the sake of b*. There seem to be at least two ways in which *S* might do *a for the sake of b*. *S* might have her teeth drilled in a painful manner for the sake of dental health. In such a case, the goal is extrinsic to the action done for its sake. On the other hand, *S* might go to the opera, have a nice post-opera dinner, and spend the next day visiting a grand cathedral all for the sake of having a nice vacation. When he does these things, S pursues them for the sake not of something extrinsic to the actions themselves; the activities he pursues are on the contrary partially *constitutive* of a good vacation.

Given that some means are constitutive of the ends to which they are

LIVING WELL

means, it remains open that Aristotle is thinking of a range of human goods done both for the sake of happiness and as desired in themselves. Presumably, given Aristotle's unmistakable emphasis on the centrality of rationality in his characterization of the human good, one must expect that any collection of constitutive means will perforce at a minimum be a well-structured expression of intellectual virtue, rather than an assorted motley jumbled together with no internal order. If that is so, one may read Aristotle's conception of the human good as both intellectual and encompassing: intellectual in the sense that it gives pride of place to contemplation and encompassing in that non-contemplative virtuous activity will display a rationally balanced structure, one likely resulting from deliberation regarding the optimal form of life for creatures with features of the sorts human beings manifest essentially.

Of course, these initial suggestions are intended to open rather than close a central controversy surrounding the theory of human happiness propounded in Aristotle's *Nicomachean Ethics*. When investigating these matters further, it serves to reflect upon a sometimes unduly neglected aspect of the theory Aristotle develops in that work, namely that its account of human goodness cannot be shorn from the metaphysical psychology undergirding it. The question of human happiness, as Aristotle understands it, is a question *about human beings*, and is accordingly a question whose answer must be rooted in facts about such beings, including centrally the fact that humans are intentional agents acting for ends. It emerges from his essentialism that human ends are not chosen by human whim, but given by human nature. Consequently, Aristotle concludes, those seeking happiness discover rather than concoct their ends; when they do, they may order their actions rightly, that is, towards the actualization of their specifically human capacities.

8.8 Conclusions

When compared with his other less user-friendly works, Aristotle's *Nicomachean Ethics* may appear relatively accessible and non-technical. In some respects, this appearance is accurate. The work is not so heavily replete with Aristotle's characteristic terminology as are some of his other more technical works. Moreover, in part because it is informed by the close observation of actual moral psychologies, some passages in the *Nicomachean Ethics* resonate readily with our own observations of the virtuous and the vicious. To some extent, the

LIVING WELL

non-technical character of the work reflects Aristotle's own stated judgment that undue precision is inappropriate to ethics, since excessive exactness imposes a demand on the human sciences which is more appropriate only to other more austere and abstract enterprises, like mathematics (*EN* 1094b11–14, 1098a28–34). The study of ethics must be responsive to the contingent vagaries flowing through human conduct; to expect the production of precise formulas suitable to every possible circumstance will dispose us to indulge in idle digressions incapable of providing us with the action-guiding principles we seek.

So much acknowledged, it must also be said that in many more important ways the appearance of accessibility and non-technicality in the *Nicomachean Ethics* is deceptive and misleading. Although he does not pause to attract attention to the fact, Aristotle's ethical theory draws heavily upon his metaphysical and psychological theories. Because he is interested in the best life *for human beings*, he takes it for granted this will be the life of those beings whose essences and natures he has already explored and characterized elsewhere. Indeed, in the first instance, Aristotle's ethical theory presupposes that human beings have an essence of a determinate and stable sort, and that consequently when it comes time to determine what is best for such creatures it will be necessary to advert to their core, essential features. This is why Aristotle does not feel the need to inveigh at length against subjectivist conceptions of happiness: since we are talking about the good for humans, and humans are a certain way by nature, those who suppose that happiness consists in simple desire satisfaction will have failed to come to terms with a central and inescapable fact about desire: people can and do desire things which are bad for them, with the sad result that people can and do live suboptimal lives. These are, then, lives they would really rather not be leading, lives they would not have desired had they fully apprehended how best to pursue their own well-being.

Looked at from this perspective, Aristotle's celebrated function argument is both less ambitious and more successful than is sometimes supposed. He does not presume that by this argument he can prove that human beings have a determinate nature, a specifiable function, and a characteristic good. Rather, he seeks mainly in this argument to *identify* the human function he has elsewhere analysed and thereby to characterize that good which is best for human beings. This good, he argues, will be one which is good in itself, good for nothing beyond itself, complete, and such that its presence will make a life lacking in nothing. Such a good we may term *eudaimonia* – happiness or human flourishing. Without

LIVING WELL

explication, however, no such term is terribly informative. We all say that we want happiness. If we disagree about what happiness consists in, then our verbal agreement merely masks other deep and important disagreements about life's most precious prize. If we accept an objective conception of happiness rooted in features of the human essence, then it makes sense to inquire, as Aristotle inquires, into those human features whose best expression yields the optimal sort of life available to us.

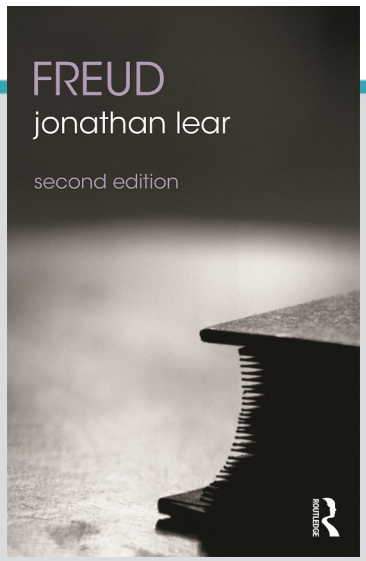
Aristotle takes it as obvious, almost beyond question, that each of us desires the best life we can secure for ourselves. Accordingly, once we have moved beyond the facile thought that the best life is whatever we happen to suppose it to be, then inquiries into human virtue (or excellence; *aretê*) of the sort engaged by Aristotle in the *Nicomachean Ethics* make perfect sense, and are, for the reflexively enlightened, almost inevitable. After all, suggests Aristotle, if we want what is good for ourselves, what is really good and not merely what happens to appeal to the whim of the moment, then it behoves us to explore what that good might be. Any such exploration will take us outside of our current subjective preferences, which may be enlightened or may be benighted, and into a consideration of the character of human excellence.

Since such excellence is trivially the excellence of human beings, we would be wise to begin our inquiry into the human good with a clear-headed conception of the character of human nature. In pursuing this inquiry, Aristotle presupposes an essentialist framework articulated within his four-causal explanatory schema, with its ineliminably teleological components. Although he does little to argue for this framework within the *Nicomachean Ethics*, Aristotle plainly presupposes familiarity with its basic precepts when advancing this brand of virtue ethics. For this reason, the *Nicomachean Ethics* is, so to speak, surreptitiously technical. Consequently, an eventual appraisal of Aristotle's ethical theory will equally implicate the sympathetic critic in a consideration of the psychological and metaphysical theories underpinning and informing it. To the degree that those theories are defensible, Aristotle's ethical theory will have much to commend it. By the same token, where those theories fail to withstand criticism, they may tend to leave Aristotle's ethical *eudaimonism* stranded, in search of the moorings without which it will be best left unembraced.

CHAPTER

3

MORALITY AND RELIGION



MORALITY AND RELIGION

Excerpted from *Freud, 2nd Edition*

1. The case against morality

Freud is famous for offering a psychoanalytic critique of morality and religious belief. It is as though he puts Western civilization itself on the couch. His aim is 'to make the unconscious conscious' – that is, he wants to show that morality and religious belief have different origins and serve different purposes than they claim. These are grand reflections about the meaning of civilization. In my opinion, they are – or, more accurately, they *have become* – the least valuable aspect of Freud's work. Perhaps in their time they served as a moment of critique. They do show ways that people can make use of moral or religious systems to gratify unconscious needs. But it is quite a stretch from that claim to the claim that this is the hidden meaning of religion and morality. To justify this latter claim, Freud would need an argument that the possibilities he uncovers are all the possibilities there are when it comes to morality and religion. Freud gives no such argument. Rather, he puts forward two paradigms –and invites readers to join him in thinking that this is all morality and religion amount to. The problem is not just a flaw in Freud's argument. There have been terrible human costs in going along with him. For generations, psychoanalytic institutes refused to train people who admitted to religious belief, on the grounds that they were fixated on unresolved infantile wishes. We do not know how many religiously oriented people – who might otherwise have benefited from psychoanalytic treatment – stayed away because they feared that analysts would try to talk them out of their commitments. We have reports from analysts that analysands found it difficult to talk about their religious beliefs, assuming ahead of time that their analysts must be atheists. They also report that analysands who were atheists assumed a kind of 'knowing alliance' with the analyst – and one can only wonder how often that went unanalyzed. But the harm is not just what these individuals have suffered. As a result, the psychoanalytic profession deprived itself of a nuanced understanding of what the analysis of religiously committed individuals might look like. In a similar vein, it deprived itself of an opportunity to contribute to a robust conception of a flourishing ethical life – because it assumed that morality must be a system of repression and discontent. The aim of this chapter is ground-clearing.

In introducing the reader to Freud's critiques of morality and religion, I will show the limits of their validity. Seeing how Freud'sarguments fall short will, I hope, open up possibilities for a deeper psychoanalytic understanding of the meanings of moral and ethical commitments in human life. This ought to make possible a more robust moral psychology.

MORALITY AND RELIGION

Excerpted from *Freud, 2nd Edition*

Freud's critiques of morality and religious belief have the form of a genealogy. In general, genealogies are stories of origins that are meant to have evaluative force. There are two dimensions along which a genealogy can be classified. First, genealogies can be either *legitimating* or *delegitimating* in intent. That is, a genealogy can seek either to valorize or to undermine via its account of how something comes to be. Second, the account can be broadly *naturalistic* or *supernatural*. Either it limits itself to an account of how something could come to be as a phenomenon of nature; or it draws on a source transcending nature as part of the account of origin. In principle, a legitimating genealogy could be either naturalistic or supernatural, and similarly for a de-legitimating one. But the original genealogies tended to be legitimating and supernatural. So, for example, the first recorded use in the *Oxford English Dictionary* is from 1300: 'Tuix Abraham and king daui, Yee herken nov be geneologi.' This genealogy is intended as a pedigree that reveals divine sanction. It valorizes Daui, and legitimates his reign, by claiming that he descends from Abraham. And Abraham is chosen by God.

By contrast, Freud's argument claims to be naturalistic and delegitimating: if we come to understand how morality arises as a natural phenomenon – as a set of institutions and practices in which human beings come to participate – we shall see that its own claims to legitimacy are false. Even worse, we shall discover thatmorality's actual aims run counter to its purported aims, and that morality is actually inimical to human well-being. As was said at the beginning of this book, Freud was not a philosopher. He seems ignorant of the ancient Greek approaches to ethics, in which the virtues – or excellences of character – are seen as contributing to a happy life. And although he does mention Kant's categorical imperative, he is not concerned with its place in the overall Kantian approach to practical reason. It is cited more as a moral dictum, along the lines of the golden rule. Freud is concerned with morality as it is lived in society – or, as it was lived in early twentieth-century Europe. These were a normatively governed set of practices and understandings of how one ought to behave with respect to others. Insofar as justification was invoked, it was by appeal to the Ten Commandments and the teachings of Jesus – in particular, his teaching to love thy neighbor as thyself. In Freud's view, society's justification for its moral practices is a legitimating, supernatural genealogy. In response, Freud is going to offer a naturalistic, delegitimating genealogy of those same practices. Freud's account of the rise of a moral capacity in humans is broadly Darwinian in structure: he gives an account of how the moral capacity comes to be selected in humans. Such an

MORALITY AND RELIGION

account shows how a phenomenon– such as the capacity for morality – can arise even though no one chose or designed it.

As we saw in the last chapter, Freud thinks the human capacity for morality arises largely as a solution to the problem of aggression.On the one hand, aggression has been selected in humans: our non-aggressive predecessors tended to get killed off before they could reproduce. On the other hand, if humans were merely aggressive animals, they would kill but they would also be under constant threat of being killed. A better solution to the problem of survival is that humans should be able to form societies that can protect their members from the aggression of other societies as well as from the menaces of nature. Society thus needs to be a way of minimizing the aggressive impulses of members of society against each other. So far, Freud's genealogy is similar to various accounts that have been given in the philosophical tradition. What makes Freud's case distinctive is his account of how human aggression is deployed in the service of curbing aggression. For what happens to the inhibited aggression?

> Something very remarkable, which we should never have guessed and which is nevertheless quite obvious. His aggressiveness is introjected, internalized; it is in point of fact, sent back to where it came from – that is, directed toward his own ego.There it is taken over by a portion of the ego which sets itself over against the rest of the ego as superego, and which now in the form of conscience, is ready to put into action against the ego the same harsh aggressiveness that the ego would have liked to satisfy upon other extraneous individuals.

But how does civilization *do* this? After all, civilization is not itself an actor, a participant in history. Freud needs a naturalistic account of how this 'achievement' of civilization comes to be. Only then will Freud distinguish himself from Nietzsche. In *On the genealogy of morality*, Nietzsche argues that guilt and bad conscience result from human aggression turned in on the self. But he gives no account of how this transformation occurs. Freud, by contrast, tries to workout a dynamic psychological account of how this inversion of aggression comes about. His account is derived from his clinical work with patients, and it has two aspects.

The first aspect concerns socialization within the family. As we saw in our discussion of the Oedipus complex in the last chapter,the process of a child entering the family necessarily involves some turning inward of aggression. Here Freud describes it at the right level of generality:

MORALITY AND RELIGION

> A considerable amount of aggressiveness must be developed in the child against the authority which prevents him from having his first, but none the less most important, satisfactions ... but he is obliged to renounce the satisfaction of this revengeful aggressiveness. He finds his way out of this economically difficult situation with the help of familiar mechanisms. By means of identification he takes the unattackable authority into himself. The authority now turns into his superego and enters into possession of all the aggressiveness which a child would have liked to exercise against it.

This is a strategic outcome no one planned. As we have seen, these are basic mental tropisms that are not themselves the outcome of choice. Once lodged inside, the superego figure becomes a vehicle through which the child's own aggression is now turned on herself.This configuration is selected because it is socially advantageous.

The second aspect is the social institution of morality. Morality provides a cultural vehicle by which the psychic transformation of the child is reinforced and given a particular cultural form. Morality functions as a 'cultural superego': it provides an explicit and shared set of practices, customs and rules that bind the members of society together both socially and psychically. These are rules that can be internalized; and as such they come to form part of the adult's superego. The individual members of society are bound together in part because each person's superego has been shaped according to a common cultural template. This is the psychic precipitate of morality. According to Freud, morality is basically a set of cultural practices and precepts that takes hold of the natural vicissitudes of the Oedipus complex and turns them to society's advantage.

This account of how morality comes to take hold has a number of significant consequences – none of them pleasant for the individual:

- Morality makes us unhappy.

The idea that morality promotes human happiness or fulfillment is,Freud thinks, exposed as wishful fantasy. For the psychic structure that morality fosters is a structure of individual human suffering: a punishing superego is set over against an inhibited ego. Outwardly and consciously, the person may be an upstanding member of society. Inwardly, and perhaps unconsciously, he is inhibited from pursuing his desires; and thus lives in frustration. Virtue is not its own reward. Indeed, the moral life, according to Freud,is necessarily and constitutionally a life of suffering. 'The two processes of individual and cultural development must stand

MORALITY AND RELIGION

in hostile opposition to each other, and mutually dispute the ground.'

- The relation between individual and society is necessarily unharmonious.

Freud understood that this insight was unsettling because it provided a blow to 'the naive self-love of men' – the sense of their deserved place in the world. Society is not there to serve human purposes, humans are there to serve society – often at significant psychic cost to themselves. A person may carve out a life for herself in society; but the essence of morality, Freud thinks, is constraint and prohibition. Morality is the institution that distorts individual human well-being for the sake of civilization.

- Morality facilitates a special kind of viciousness.

The prohibitions of morality are not just demands of society; they are internalized and become prohibitions of a person's own superego. Since the superego is sensitive to a person's thoughts as well as deeds, there is no place to hide. A person's wishful and aggressive thoughts will inevitably contradict the prohibitions of society. And thus people will inevitably incur the wrath of their own superegos. Guilt is thus an inevitable condition of living in civilization. And given the way in which the superego enables people to turn their aggression onto themselves, a truly terrifying economy is established:

> here at last comes an idea which belongs entirely to psychoanalysis and which is foreign to people's ordinary way of thinking. This idea is of a sort which enables us to understand why the subject-matter was bound to seem so confused and obscure to us. For it tells us that conscience (or more correctly the anxiety that becomes conscience) is indeed the cause of instinctual renunciation to begin with, but that later the relationship is reversed. Every renunciation of instinct now becomes a dynamic source of conscience and every fresh renunciation increases the latter's severity and intolerance.

This is a kind of sorcerer's apprentice of moral asceticism: the more 'moral' one becomes, the more aggression is inhibited from discharge in the social world, and thus it is turned inward on oneself. There arises the furiously moral person – the 'saint' – who takes himself to be such a sinner. Freud thinks that such a person has a basically correct assessment of his internal situation. So too arises the phenomenon Freud called moral masochism: the person perversely dedicated to castigating himself for being so awful.

MORALITY AND RELIGION

2. The morality system

Freud has, I think, given a psychodynamic account of what the philosopher Bernard Williams called 'the morality system.' This is 'a *special* system, a *particular variety* of ethical thought.' Williams thinks that in coming to understand it we will at the same time see 'why we would be better off without it.' Note that the critique is not about moral life *per se*, but about an unhealthy yet ultimately optional distortion. The morality system, Williams argues, takes our ordinary notion of obligation, which on its own has important uses, and turns it into a special, highly charged notion of *moral* obligation – to which it attaches too much significance.

> Moral obligation is inescapable. ... Once I am under the obligation there is no escaping it, and the fact that a given agent would prefer not to be in this system or bound by its rules will not excuse him; nor will blaming him be based on a misunderstanding.Blame is the characteristic reaction of the morality system. The remorse or self-reproach or guilt ... is the characteristic first-personal reaction within the system, and if an agent never felt such sentiments, he would not belong to the morality system or be a full moral agent in its terms.

The morality system, Williams tells us, tries to turn everything into an obligation and it 'encourages the idea, only an obligation can beat an obligation.'

In morality so conceived, we have a closed system of blame and guilt with no upper bound on stringency and no escape from obligation.Freud gives the psychological account of how this system can take hold of us. But he also raises a problem for Williams about the difficulty of a way out. Williams says: 'In order to see around the intimidating structure that morality has made out of the idea of obligation, we need an account of what obligations are when they are rightly seen as merely one kind of ethical consideration among others. This account will help to lead us away from morality's special notion of moral obligation, and eventually out of the morality system altogether.' From a Freudian perspective, Williams is being too optimistic about the power of a thoughtfully reasoned 'account'to lead us out of the morality system. If Freud is right about the psychodynamics by which the morality system takes hold of us, then a reasoned account on its own will not do the trick. We need a psychodynamic account about how, over time, the morality system might loosen its grip. What would be required for the superego to give up its punishing stance of guilt and obligation? What cultural developments might facilitate this process in terms of large-scale cultural shifts? And is there a

MORALITY AND RELIGION

particular form of political critique that would be appropriate – one that used psychoanalytic insight to offer alternatives to political societies that make undue use of the morality system to organize and subdue its citizens? These are the kind of questions an informed psychoanalysis might help us answer. Freud had a marvelous ability to turn a problem – for instance, transference– into a solution. If the problem of the morality system is a punishing superego, why isn't the solution the conditions of healthy relations between ego and superego? This would be a natural place for psychoanalysis to make a lasting contribution to moral psychology.In effect, this is what would be involved in taking up the inheritance of an Aristotelian approach. For Aristotle, as for Plato before him,happiness required psychic harmony between the rational and non-rational parts of the soul – and such harmony was a condition of an ethically virtuous life. If Freud had been able to make the same kind of move as he made with transference, he would have gone from intrapsychic relations as a problem for moral life, to intrapsychic relations as the conditions for a satisfying moral life.

3. Pleasure versus happiness

I suspect there is a contingent reason he did not go down this route: in his theoretical research he focused on pleasure, rather than on happiness. His concern with pleasure goes back to the beginning of his career, long before he had a structural theory of the mind – and thus before he could formulate an Aristotelian approach to moral psychology. What attracted him to the phenomenon of pleasure is that he thought he could give a quasi-mechanistic account of it. It fit his conception of a naturalistic project by which one could plausibly show how complex workings of the mind could be built up from elementary operations of a 'psychic mechanism.' The aim of the psychic mechanism, as he conceived it, was to discharge pent-up psychic energy, and this discharge was experienced as pleasure. This schematic picture has heuristic value: humans do live under psychic pressure, and Freud was a master at charting the myriad ways they seek release. But this picture also blinkered Freud's thinking. For if built-up tension is in itself unpleasurable, it becomes easy to assume that such tension is itself a condition of unhappiness. If one holds onto this theory of pleasure-as-discharge, the subsequent discovery of psychological structure will not seem like an occasion to rethink the possibilities of human happiness. The ego and superego are variously in the business of inhibiting, reshaping, redirecting the wishes of the id. But if the gratification of these wishes would provide the most immediate and direct

MORALITY AND RELIGION

discharge, it should provide the most pleasure. According to this mechanistic theory, then, psychological structure itself inhibits pleasure and is thus an occasion for unpleasure. Again it becomes easy to assume that psychic structure inevitably causes unhappiness. This would seem to be confirmed by Freud's clinical realization that a punishing superego was the key to the individual'sdiscontent in civilization. Since the human condition would then appear to be inevitably bound up with unhappiness, the only real questions would be how to minimize it and fend off pathological distortions. There would not be room for the question of a psychologically harmonious, happy ethical life to arise. None of this reasoning is ultimately justified, but one can see how it hangs together. There were disruptions in Freud's thought that could have been the basis for a thorough reconceptualization of his theory of pleasure. By the time he wrote 'The economic problem of masochism'(1924), Freud realized that increases in tension can be pleasurable. This recognition could have been an occasion to consider how certain forms of tension-filled dynamic psychological structure could in themselves be pleasurable. For whatever reason, Freud did not take up this opportunity, and thus he was not in a position to use the discovery of psychological structure to think afresh about the possibilities for human happiness.

4. Critique of religious belief

Freud argued that religious belief is illusion. He meant this in a precise sense: a belief is an *illusion* if it is derived from human wishes. Illusions are by their very nature misleading. For people take their beliefs to be responsive to the way things are. So if a belief is held in place by wishes, people are misled about their orientation to the world. Beliefs can be true or false; the same holds for illusions. It is not out of the question for an illusion to be true.The essential problem for an illusion, then, is that we are mistaken about the basis of our commitment to it. We take it to be a belief based on responsiveness to the world; in fact, it is held in place by primordial wishes of which we are unconscious.

Freud's argument is oblique. He does not address religion directly; and ostensibly he makes no claims about whether religious beliefs are true or false. His claim is rather that religious beliefs are illusions. That is, whatever the truth of religious claims, the fact that we believe them is not based on that truth, but rather on infantile wishes. His expectation seems to be that once we recognize these beliefs as illusions, and come to see the kind of wishes they gratify, the temptation towards religious belief will fall away. At the very least, we will see that we *ought*

MORALITY AND RELIGION

to give up religious belief.

His argument is flawed. But before looking for the flaw, it is worth noting that Freud's aim is more than the dissolution of religious belief. He is also attacking what he takes to be the foundation of morality. Morality, he thinks, depends on its claim to be carrying out the teachings of the Hebrew and Christian Bibles. The ultimate authority for morality is the Word of God. Thus morality, on Freud's understanding, provides a genealogical defense of its authority that is absolute and supernatural. In response, we can now see, Freud offers a counter-genealogy that is meant to be de-legitimating and naturalistic. We have already seen the first stage of Freud's argument by which he offers an alternative, deflationary account of how the moral capacity arises in people. We are now at the second stage in which Freud seeks to undermine morality's appeal to a religious foundation. Religious belief, Freud argues, arises from an infantile prototype: our earliest experiences of helplessness. Religion emerges as a cultural elaboration of childhood fantasies whose function is to protect us against a sense of utter vulnerability. In response, Freud says, we wishfully imagine that the world is ordered according to a higher purpose and we each have a proper role within it.

> Over each one of us there watches a benevolent Providence which is only seemingly stern and which will not suffer us to become a plaything of the over-mighty and pitiless force of nature. Death itself is not extinction, is not a return to inorganic lifelessness, but the beginning of a new kind of existence which lies on the path of development to something higher.

Freud diagnoses this as a manifestation in adult life of an infantile longing for the father – a wish for a powerful, protective figure. This is why Freud thinks religion is an illusion: it is held in place by 'the oldest, strongest and most urgent wishes of mankind. The secret of their strength lies in the strength of those wishes.' And he is scathing in his judgment: 'The whole thing is so patently infantile, so foreign to reality, that to anyone with a friendly attitude to humanity it is painful to think that the great majority of mortals will never be able to rise above this view of life.' Even more pathetic, Freud thinks, are those educated people who ought to know better, but still try to defend religion 'in a series of pitiful rearguard actions.' I suspect that would be Freud's charge against the author of this chapter.

That said, there is a problem in Freud's argument. In a nutshell, it contains its own kernel of wishfulness. And this wishfulness compromises his argument. It is helpful to ask how Freud's argument is meant to persuade. That is, what is its rhetorical strategy? If we take the idea of illusion seriously, there is a question of

MORALITY AND RELIGION

how we could come to recognize any of our beliefs as illusions. An illusion does not seem to be an illusion to those who are in its grip. And if illusions are held in place by primordial wishes, one would expect them to be tenacious. In particular, one should not expect people to recognize their illusions simply by being told by another that their purported beliefs are illusory. So, if we are suffering from illusion,how will it help us for Freud simply to tell us that we are?

Obviously, if Freud's intended readers are only other non-believers,then there is no need to persuade them. It is easy enough for author and readers to agree that *other people* suffer from illusion. Illusion tends to be the kind of thing other people suffer from.

This would be the atheist version of preaching to the choir. The more interesting case is to think that Freud is also writing for religious believers, as well as for agnostics and those who are suspended in a limbo between belief and non-belief. How, one might ask, is Freud's diagnosis meant to reach such readers? Even if one accepts Freud's diagnosis that religious belief has a wishful component, this need not on its own give a person reason to abandon religious commitment. It is a longstanding belief among religious thinkers that, precisely because we come into the world as children, religion needs a childish – that is, age-appropriate –component. From such a religious perspective, this is the point of many myths and stories. It is thus not a criticism, from this perspective, to point out that there is an infantile dimension to religious belief. The crucial question would be: what, religiously speaking, are the possibilities for growing up? Perhaps the wishful illusions of childhood were no more than steps on a spiritual ladder that was meant to be kicked away at a later stage. To eliminate this possibility, Freud would need to show that there is nothing to religious belief *other* than illusion. He does not try to do this. For whom then might Freud's critique function *as* a critique? The argument seems designed to appeal to three broad groups: agnostics, atheists and people who are merely going through the motions of religious rituals. Perhaps they are lazy; perhaps they are sitting on the fence. For agnostics and atheists, Freud's argument would likely seem to confirm their doubts; for the person who has been going through the motions, it might tip her into viewing her own behavior as childish. It might be a step along a path of disenchantment. We have examined a case where Freud's argument fails to persuade, and cases where it would. What we can see is this: whether the argument does or does not persuade does not depend on accepting Freud's premise that religious belief has a wishful component. All sides we have

MORALITY AND RELIGION

considered accept that. For Freud's argument to have the rhetorical force he intended, one needs to consider the context in which he took himself to be writing. Freud took himself to be writing within an historical epoch of secularization within Europe, and he intended his critique to further that process. His argument is directed at those who have *already* lost their religious belief or those who are *already* wavering in agnosticism or those who are *already* participating in the social rituals of religion in a weak way. For such a reader, the argument may facilitate their journey towards a non-religious life. In this way, Freud takes his argument to be helping history along. But what is Freud's view of history?

5. The illusion of a future

Freud thinks we have reached an historic epoch in which we can simply see that his analysis is true. If we look to individual development, Freud says, we see that a person develops through psychological stages. In particular, the inevitable Oedipal crisis of childhood is eventually outgrown, and falls away. 'In just the same way,' he says, 'one might assume, humanity as a whole, in its development through the ages, fell into states analogous to the neuroses, and for the same reasons.'

> Religion would thus be the universal obsessional neurosis of humanity; like the obsessional neurosis of children, it arose out of the Oedipus complex, out of the relation to the father.If this view is right, it is to be supposed that a turning-away from religion is bound to occur with the fatal inevitability of a process of growth and that we find ourselves at this very juncture in the middle of that phase of development.

On Freud's account, then, his interpretation will be persuasive because we are historically ready to face the truth. But just how wishful is such an image of historical progress? Freud gives us no reason to believe the history of civilization proceeds 'in just the same way' as the development of an individual out of childhood.And, of course, there are reasons to reject this picture as a progressivist fantasy. But Freud was unwilling to countenance any other explanatory hypotheses for this movement other than an inevitably emerging truth of which he is an avatar. But if we look to the argument, a comparison with the Danish philosopher Søren Kierkegaard is illuminating. Kierkegaard thought that Christendom was a 'dreadful illusion.' He did not, like Freud, have an explicit and technical definition of illusion. We can nevertheless see an important area of agreement with Freud.

> Christendom, for Kierkegaard, referred to the totality of social practices,

MORALITY AND RELIGION

customs and rituals of his day that were accepted as an expression of Christian faith. To say that Christendom was an illusion was, for Kierkegaard, to say that these social practices trapped one in a misleading fantasy of religious commitment. But, for Kierkegaard, this recognition was meant to be preparatory to a more genuine religious engagement. To do this, however, one would have to buck the trends of age. The illusion of Christendom was itself part of a general decadence of the historical epoch. The modern age had produced changes in mass communication – in particular, the capacity to publish newspapers and pamphlets – and increased production in consumer goods – and this led to what Kierkegaard called the leveling of the age. Newspapers tell one what one should believe – although they are doing little more than passing along unsubstantiated gossip – and advertisements tell one what one should want. This, Kierkegaard thought, led to the collapse of the individual (in particular, an individual's sense of responsibility) and in its place a 'crowd' formed, governed by rumor and fashion. In a crowd, genuine religious engagement becomes impossible. This is obviously a thumbnail sketch, but it is enough to see that while Freud and Kierkegaard start from the same phenomenon – the decline of religious conviction in contemporary Europe – and they both believe that common religious practice is illusion, they draw opposite conclusions.

Once one sees that Freud's argument need not push one in only one direction, it is easier to see the conclusion Freud does draw as wishfully heroic. He imagines a conversation with a religious interlocutor:

> We desire the same things, but you are more impatient, more exacting, and – why should I not say it? – more self-seeking than I and those on my side. You would have the state of bliss begin directly after death; you expect the impossible from it and you will not surrender the claims of the individual. Our God, *Λογος* [Logos], will fulfill whichever of these wishes nature outside us allows, but he will do it very gradually, only in the unforeseeable future, and for a new generation of men. He promises no compensation for us, who suffer grievously from life. On the way to this distant goal your religious doctrines will have to be discarded, no matter whether the first attempts fail,or whether the first substitutes prove to be untenable. You know why: in the long run nothing can withstand reason and experience, and the contradiction which religion offers to both is all too palpable. Even purified religious ideas cannot escape this fate, so long as they try to preserve anything of the consolation of religion.

MORALITY AND RELIGION

Why should Freud of all people believe this? Has he not taught us that we are always subject to wishful and aggressive fantasies of which we are largely unaware? If so, should one not expect that whatever genuine achievements of reason and experience occur,they will tend to be fragile and subject to reversal? Whatever one thinks about religion and religious belief, one should by now see that this is less the future of an illusion, than the illusion of a future. History is assumed to be progressive, inevitable and truth revealing.This is a triumphal story of human progress in which one can play a decisive role if only one is brave enough to face the emerging truth. Obviously, there is no reason to go along with this Enlightenment fantasy about the significance of human history. So,in the name of analyzing the fantasy underlying religious belief, Freud participated in his own fantasy of inevitable historical progress, which included secularization as a hallmark of that progress. There is reason to think that this closed down Freud's curiosity: he was disposed to see religious commitment as historically retrogressive. If he could find a kernel of wishfulness in that commitment that was sufficient; it was as though there was nothing more to look for. As a result, Freud blinded himself to the possible complexity of religious belief. We are still living in Freud's shadow. Instead of assuming ahead of time that religious commitment must be illusion, we need to know more than we now do about the place of religious belief in analysands' lives. This is a place where psychoanalysis could make a contribution: helping us to understand in robust detail the myriad places of religious commitment in individuals' lives. We get in the way of that project by assuming ahead of time that there is only one such place, and we already know what it is.

6. Primal crime

Freud also had an illusion of the past. He claimed that history as we know it was inaugurated by a murder of the 'primal father' by the brothers and sons. In prehistory, humans were organized in hordes. Each horde was ruled by a primal father who subjugated the other men, and had sexual access to all the women. Eventually the men banded together and killed the primal father; and society was organized around an agreement that no one else would take his place. Instead, the incest taboo was established,which facilitated the sharing of women among the men, and social relations between them. But this primal crime also laid down an archaic heritage of guilt. For the father who was hated was also the father who was loved. The murder was so traumatic that permanent memory trace was laid

MORALITY AND RELIGION

down in the human race –a phylogenetic inheritance as Freud called it. (Genes had not yet been discovered, but I suspect Freud would have been happy with the idea that the murder made a genetic difference; and this genetic alteration continues to be passed down through the generations.) At the end of his career, Freud makes a further astonishing claim: that the Jewish people murdered Moses as a repetition and recreation of that primal crime. According to Freud, Moses was an Egyptian who tried to impose monotheism on the recalcitrant Hebrews. It is, says Freud, the murder of Moses that provokes the wishful fantasy of the return of the Messiah. And it is the cover-up of this murder that results in official Judaism taking on the form of an obsessional neurosis: structured by endless rituals of cleanliness and purification. Moreover, Christ's coming and his killing was another repetition and re-creation of the primal crime:

> If Moses was the first Messiah, Christ became his substitute and successor, and Paul could exclaim to the peoples with some historical justification: 'Look the Messiah has really come: he has been murdered before your eyes!' Then too there is a real piece of historical truth in Christ's resurrection, for he was the resurrected Moses and behind him the returned primal father of the primitive horde, transfigured and, as the son, put in the place of the father. The poor Jewish people, who with their habitual stubbornness continued to disavow the father's murder, atoned heavily for it in the course of time. They were constantly met with the reproach 'You killed our God!' And this reproach is true, if it is correctly translated. If it is brought into relation with the history of religions it runs: 'you will not admit that you murdered God (the primal picture of God, the primal father, and his later reincarnations).' There should be an addition declaring: 'We did the same thing, to be sure, but we have admitted it and since then we have been absolved.'

These are extraordinary claims – and they are based on almost no evidence. Freud admits that, when it comes to a memory-trace of an ancient crime, he has nothing more to go on than the tenacity of Oedipal fantasies in the psychoanalytic situation. He does not think that the power of guilt, ambivalence and aggression that he sees in his neurotic patients can be explained on the basis of their imaginative and emotional life alone, nor on the basis of their experiences with others. Similarly, he did not think it possible for religion to be transmitted from generation to generation using only cultural and psychological means. To explain the tenacity with which the Jews held onto their religion it was not enough, Freud

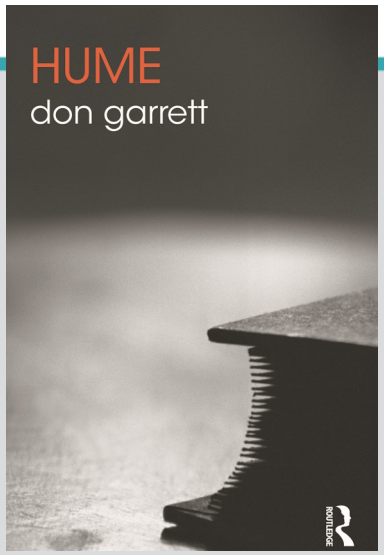
MORALITY AND RELIGION

thought, to cite Jewish rituals, festivals and teachings. Nor was it enough to include an account of how wishful infantile fantasies get entwined in religious myths. In Freud's opinion, these facts alone could not explain the stubborn persistence of Judaism. There had to be an actual crime whose trace was laid down in human memory, a primal murder that was repeated by the Jews, but never acknowledged. Freud is making a bold assertion, but there is really no basis for it. And if we consider the place of this speculation in the larger framework of his thought, Freud is in effect attacking his own life's work. He has spent his career showing the power of unconscious fantasy to shape a life, but when it comes to our religious lives, he claims this cannot be explained by the power of human imagination, culture and rituals alone. He is talking particularly about Judaism and Christianity: religions in which God intervenes in history and interacts with specific human individuals. Freud agrees with the religions to this extent: for these religions to be possible there must have been a significant actual historical event. These religions cannot, he thinks, be understood simply as a product of the human imagination. But he takes that actual event to be secular: the murder of the primal father, followed by subsequent re-enactments with Moses and then with Jesus. If Freud's argument had been sound, he would have given a thoroughly naturalist account of religious experience. Obviously, it is in principle possible to give a naturalist account only invoking human imagination and culture. But by invoking an actual event, Freud thought he had uncovered the hidden meaning of these fantasies. He thought he had given a secular and naturalist counterpart to original sin (the primal crime) and to the transmission of hereditary sin (phylogenetic inheritance). Without the actual crime, there would always be a question of why human imagination and culture took this form rather than some other –and there would be no place to look other than further delving into imagination and culture. Freud wants the primal crime to serve as an Archimedean point. But this isn't an Archimedean point; it's a fantasy of having achieved one. In effect, Freud constructs his own myth of origins. He hides this from himself by cloaking his myth in the garb of a naturalistic account of human development. Freudian psychoanalysis aims to be a naturalistic account of human mental life – an account that includes culture,rituals and social institutions. It is within this important project that Freud embeds his own illusion: that science will answer the fundamental question of how to live. So, he thinks that any rational person who accepts the findings of science ought to give up religious belief. This is a wishful illusion of rationality and scientific progress. It captured Freud's imagination, but there is no reason for us to be in its thrall.

CHAPTER

3

MORALITY AND VIRTUE



MORALITY AND VIRTUE

Excerpted from *Hume*

Hume's moral philosophy is a distinctive and influential combination of *virtue ethics* and *moral sense theory* (or *sentimentalism*). For him, VIRTUE and VICE are the fundamental normative concepts structuring the moral domain of value through their application to character, while the concepts used in the moral evaluation of actions derive their normativity from these [5.1]. In this he stands with the ancient moralists he admired and in opposition to those who treat the evaluation of actions for conformity to duty as fundamental. In particular, he stands in substantial opposition to the long tradition of *natural law* tracing back at least to Thomas Aquinas (1225–74) and renewed in modern Protestant Europe by Hugo Grotius (1583–1645), Samuel Pufendorf (1632– 94), and John Locke. In this tradition, the evident need for human beings to preserve themselves and to live together socially permits the derivation of moral precepts that can be known by reason, even without revelation, to have the force of moral law as commands of God.

At the same time, VIRTUE and VICE are for Hume also immediately sense-based concepts that arise from a moral sense consisting in the capacity to feel pleasurable moral approbation or painful moral disapprobation [4.1]. In this he stands with philosophers like Shaftesbury and Hutcheson who model the epistemology of moral qualities partly on that of the secondary qualities of bodies, and in opposition not only to natural law theorists, but also to philosophers like Clarke and Malebranche who model the epistemology of morality partly on that of mathematics.

Whereas Hutcheson treats the moral sense as directed primarily at actions performed from benevolence, however, Hume maintains that a wide range of mental traits and characteristics are equally capable of stimulating the moral sense. When the mind feels non-moral pleasures or pains and regards them as caused by a mental characteristic it is considering, the result is the distinctively moral pleasure or pain of moral approbation or disapprobation, respectively. Typically, though not quite always (THN 3.3.1.10/577–78; THN 3.3.1.26–29/589–90), these non-moral pleasures and pains are felt through the operation of sympathy [3.6]; and moral sentiments themselves can also be strengthened or spread to others by sympathy.

Hume draws a number of important distinctions among kinds of virtues. One way to classify them, especially prominent in the organization of the second *Enquiry*, is by means of their primary source of appeal to the moral sense as delineated in the four-part productive definition of 'virtue' or 'personal merit' as

MORALITY AND VIRTUE

mental characteristics that are useful or agreeable to their possessor or others [4.2]. Books of moral guidance of the period often drew a tripartite distinction among duties to God, duties to self, and duties to others. It would therefore have been notable that Hume omits any mention of duties to God. This omission and his positive location of the source of moral distinctions in features of distinctively human passions and taste are central to his irreligious and naturalistic philosophical purposes.

Another important distinction of kinds of virtues for Hume— especially prominent in the organization of *Treatise* Book 3—is that between "natural" and "artificial" virtues. The latter, unlike the former, depend for their existence on conventions and artifice. Within the natural virtues, those of "greatness of mind" are especially related to pride, while those of "goodness and benevolence" are particularly related to love. Still other natural virtues are among those often classified as mere "natural abilities." He distinguishes "moral obligation" from what he calls "interested" (also "self-interested" or "natural") obligation, and he argues that there is in general an interested obligation, as well as a moral obligation, to virtue. This conclusion is central to his practical aim of encouraging virtue. Like the sense-based concepts PROBABILITY, BEAUTY, and DEFORMITY, the concepts of VIRTUE and VICE require some relativization to circumstances [4.4] and allow some scope for blameless diversity [4.2].

Hume begins Book 3 of the *Treatise* by arguing, against Malebranche, Clarke, and their followers that "moral distinctions are not deriv'd from reason." This famous conclusion is a necessary preliminary to his positive account of the way in which moral distinctions are derived from a moral sense. It is also, however, one of a set of famous conclusions he offers at different points in the *Treatise* about what the faculty of reason, perhaps surprisingly, cannot do; other examples include his conclusion that reason cannot determine the inference from observed to unobserved (THN 1.3.6) [3.4, 6.1] and his conclusion that reason cannot produce the belief in the continued and distinct existence of bodies (THN 1.4.2) [3.5]. His conclusion that reason cannot be the source of moral distinctions depends, in turn, partly on an argument for yet another member of this set: his conclusion in Book 2's "Of the influencing motives of the will" that reason alone cannot determine the will to any action (THN 2.3.3) [3.7].

8.1 Reason and moral distinctions

Hume offers three arguments for the conclusion that "moral distinctions are not

MORALITY AND VIRTUE

deriv'd from reason." One of these is sometimes called the *Motivation Argument*. He is so pleased by it that he devotes a paragraph to giving it twice in a row, varying only its wording and the placement of its conclusion:

> [i] Since morals, therefore, have an influence on the actions and affections, it follows, that they cannot be deriv'd from reason; and that because reason alone, as we have already prov'd, can never have any such influence. [ii] Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason.
>
> (THN 3.1.1.6/457)

The main point of controversy, Hume recognizes, will be the premise that reason alone is impotent with respect to action, and he therefore offers to recapitulate one of his previous arguments from Book 2 for this claim in such a way as to render it "still more conclusive, and more applicable to the present subject." The previous argument he cites not the primary Impulse Argument, but rather the confirmatory Representation Argument [3.7]. His new, still-more-conclusive-and-applicable version of this argument is buttressed with the claim that not only passions but also actions and volitions are non-representational. (This is plausible for Hume because, just as passions are non-representational impressions accompanied by representational ideas, on his account, so too are volitions.) Hence, he argues, neither passions, nor actions, nor volitions can be either true or false. Yet reason, as a belief-producing faculty, is essentially a discovery, or attempted discovery, of truth and falsehood [3.3, 3.4]; passions, actions, and volitions can therefore neither conform with nor be contrary to reason, and accordingly reason alone cannot influence actions. By characterizing reason as a "discovery of truth or falshood," Hume presumably seeks to show the applicability of his argument to opponents who define 'reason' in broadly epistemic, rather than inferential, terms. In drawing his main conclusion, he is making the perhaps defensible assumption that reason as a faculty could operate to influence actions only by means of discovering (or seeming to discover) the conformity or contrariety of passions, volitions, or actions with truth—that is, their correspondence with something they represented [5.2].

Hume asserts that while this expanded version of the Representation Argument supports the key premise of the Motivation argument, it can at the same time stand alone as a second argument that moral distinctions are not derived

MORALITY AND VIRTUE

from reason, even without appeal to morality's own contrasting motivational capacity. For on the assumption that reason alone, as a belief-producing faculty, could draw a moral distinction *only* by means of discovering the conformity or contrariety of passions, volitions, or actions with truth, it also follows directly from the premises of the Representation Argument that reason alone cannot make moral distinctions.

Before proceeding to his third argument, Hume stops to consider the objection that even if passions, volitions, and actions cannot be immediately contrary to reason, they may still be so indirectly, in virtue of their causes or effects. Considering causes first, he notes two ways in which the cause of a desire or other passion leading to volition and action may be in error: (i) by including a false belief about the existence of an object, or (ii) by including a false belief about the means to some already desired end. But it is improper, he claims, to call the passions themselves contrary to reason on this account, and such errors are not, in any case, immoral or the source of moral distinctions. On the contrary, such mistakes of fact are typically morally (even if not epistemically) innocent, and if errors of this kind were the source of immorality, it seems that all such errors would give rise to immorality regardless of their object. Although it might be argued that there is a special class of "mistakes of right," their existence as *mistakes* of right would, he notes, presuppose *facts* of right as their subject matter; and these facts of right, with relevant distinctions derivable from them, would then have to have existed antecedently, without their existence being explained through any error or mistake of reason.

Turning to errors concerning effects, Hume notes that passions, volitions, and actions may themselves cause erroneous beliefs. Yet if this kind of error were the source of moral distinctions, then all production of false beliefs, even by inanimate objects, would be immoral. In addition, some immoral actions are entirely hidden and have no tendency to produce false beliefs, while the full immoral character of others is quite obvious and unmistakable. Although an act like stealing might erroneously suggest, in a way, that the object is the property of the thief, this is only because the rules of property place their "rightful" possession with another, which kind of possession is then generally assumed; in this case, therefore, any tendency to deception arises from the immorality, rather than being the source it.

In his third and final line of argument for the conclusion that moral distinctions are not derived from reason, Hume proceeds in his usual manner when

MORALITY AND VIRTUE

arguing that something is not produced by reason: namely, by arguing first that it is not caused by demonstrative reasoning and second that it is not caused by probable reasoning. If moral distinctions were made by demonstrative reasoning, he asserts, they would have to lie either in one of the four relations already identified as the bases of demonstrable knowledge strictly so-called—resemblance, contrariety, degrees in quality, and proportion in quantity [3.4]—or in some further, as yet unidentified, demonstrable relation or relations. Yet morality clearly does not lie in any of the four identified relations, and any newly proposed relation, he states, would have to meet two stringent conditions.

The first condition is that the relation should hold only between minds and external objects, and never between minds independently of external objects or between external objects independently of minds. This demand is legitimate, Hume holds, because morality applies only to circumstances that involve both minds and external objects. Many of the same relations involved in the crime of parricide also hold when an oak tree destroys its parent, for example, but the latter is not immoral. The second condition is that it should be shown how apprehension of the new relation would necessarily serve to motivate every rational being, even including the deity. This demand is legitimate, Hume holds, because it has already been established that the apprehension of moral distinctions is inherently motivating; so if reason alone is sufficient to make such distinctions, the relations it apprehends must be motivating for all possible beings possessed of reason. Yet this seems impossible to show, he argues, because the question of what considerations will motivate a given kind of being is always a causal question about the production of volitions, and causal questions can only be answered on the basis of experience. Because neither of the two required conditions can evidently be satisfied, he concludes that demonstrative reasoning cannot make moral distinctions.

Yet neither, Hume argues in a famous passage, can probable reasoning (that is, reasoning concerning matters of fact) make moral distinctions:

> Nor does this reasoning only prove, that morality consists not in any relations, that are the objects of science; but if examin'd, will prove with equal certainty, that it consists not in any matter of fact, which can be discover'd by the understanding. This is the second part of our argument; and if it can be made evident, we may conclude that morality is not an object of reason. But can there be any difficulty in proving, that vice and virtue are not matters of fact, whose existence we can infer by reason? Take

MORALITY AND VIRTUE

> any action allow'd to be vicious: Wilful murder, for instance. Examine it in all lights, and see if you can find that matter of fact, or real existence, which you call vice. In which-ever way you take it, you find only certain passions, motives, volitions, and thoughts. There is no other matter of fact in the case. The vice entirely escapes you, as long as you consider the object. You never can find it, till you turn your reflection into your own breast, and find a sentiment of disapprobation, which arises in you, towards this action. Here is a matter of fact; but it is the object of feeling, not of reason. It lies in yourself, not in the object. So that when you pronounce any action or character to be vicious, you mean nothing, but that from the constitution of your nature you have a feeling or sentiment of blame from the contemplation of it.
>
> (THN 3.1.1.26/468–69)

In order to understand this passage, it is necessary to note three important points.

First, whenever Hume writes of what someone "means" by doing something, he is describing what the person *signifies* or *gives a sign of;* he is not seeking to provide a strict semantic analysis or synonymous expression. Thus, when he writes, "when you pronounce any action or character to be vicious, you mean nothing, but that from the constitution of your nature you have a feeling or sentiment of blame from the contemplation of it," he is simply making a claim about what the act of pronouncing this indicates about the utterer; he is not proposing that a predicative use of 'vicious' is synonymous with an assertion of the existence of a causal relation between the conjunction of one's own nature with an act of contemplation, on the one hand, and a sentiment of blame on the other. Not only do Humean moral judgments not refer to the judger, but in his discussion of the correction of moral sentiments [4.1] he is quite explicit that a predicative assertion using the term 'vice' or 'virtue' can be true or correct in the absence of any actually felt sentiment (THN 3.3.1.16/582)—just as a predicative color or aesthetic assertion about an object can be true and correct even in the absence of any actual color impression or aesthetic sentiment. A moral judgment can also, of course, be false even when produced by a moral sentiment, if the sentiment would not be felt from the standard of judgment for morals.

Second, Hume is not saying that morality does not consist of matters of fact, but only that morality does not consist of matters of fact "which can be discovered by the understanding"—in other words, in this context, by reason [3.3]. In this respect, virtue, and vice are like colors and other secondary qualities; their

MORALITY AND VIRTUE

possession is a matter of fact, but one discerned in the first instance by a sense, rather than by reason. (The comma before 'which' should not distract; eighteenth-century English punctuation differs from the contemporary in freely allowing commas before restrictive clauses.)

Third, Hume's implicit location of vice "in your own breast" and his explicit identification of it as a "perception in the mind" exemplify the systematic ambiguity to which terms used to signify sense-based concepts are subject generally in Hume's Lockean usage [4.2]. He is equally emphatic, in his account of the passions in Book 2, about characterizing virtue and vice as "qualities in" the "subject" who is the person morally judged, rather than as "sentiments in" the one morally judging (THN 2.1.2.6/279; THN 2.1.7); it is only because they have this status as "qualities in a subject" that they can qualify as "causes" in the specific sense relevant to the double relation of impressions and ideas required for the origins of love and pride, hatred and humility [3.6]. In a similar way, as we have seen, he is willing to use color terms to denote qualities of bodies as well as sensations in the mind, and to use the terms "beauty" and "power" in both ways as well. Indeed, he continues the passage quoted above as follows:

> Vice and virtue, therefore, may be compar'd to sounds, colours, heat, and cold, which, according to modern philosophy, are not qualities in objects, but perceptions in the mind: and this discovery in morals, like that other in physics, is to be regarded as a considerable advancement of the speculative sciences; tho', like that too, it has little or no influence on practice. Nothing can be more real, or concern us more, than our own sentiments of pleasure and uneasiness; and if these be favourable to virtue, and unfavourable to vice, no more can be requisite to the regulation of our conduct and behaviour.
>
> (THN 3.1.1.26/469)

Thus, Hume's argument that moral distinctions cannot be made by probable reasoning alone is that such distinctions require the capacity for a distinctive kind of impression of reflection from which moral concepts originate, just as the making of color distinctions requires the capacity for a distinctive kind of impression of sensation from which color concepts originate. It is for this reason that he is willing to offer as an alternative expression of his main question this formulation: "*Whether 'tis by means of our* ideas *or* impressions *we distinguish betwixt vice and virtue, and pronounce an action blameable or praise-worthy*?" (THN 3.1.1.3/456).

MORALITY AND VIRTUE

Moral distinctions are not derived from reason, he answers in effect, because the fundamental moral concepts are instead sense-based.

In a famous paragraph that was added to the text just prior to publication, Hume concludes the first section of *Treatise* Book 3 thus:

> I cannot forbear adding to these reasonings an observation, which may, perhaps, be found of some importance. In every system of morality, which I have hitherto met with, I have always remark'd, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surpriz'd to find, that instead of the usual copulations of propositions, *is*, and *is not*, I meet with no proposition that is not connected with an *ought*, or an *ought not*. This change is imperceptible; but is, however, of the last consequence. For as this *ought*, or *ought not*, expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason shou'd be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it. But as authors do not commonly use this precaution, I shall presume to recommend it to the reader; and am perswaded, that this small attention wou'd subvert all the vulgar systems of morality, and let us see, that the distinction of vice and virtue is not founded merely on the relations of objects, nor is perceiv'd by reason.
>
> (THN 3.1.1.27/469/70)

The proposition that "one cannot derive an *ought* from an *is*" has since come to be known as *Hume's Law*. As such, it has been taken to be an important formulation of the distinction between facts and values, and discussions of its correctness have taken on something of a life of their own. The interpretation of the passage in the context of Hume's philosophy has also been a lively topic of debate in its own right, however. Some have assumed that he denies any legitimate "deduction" (that is, in Hume's usage, any inference or other derivation) of *ought* from is because he holds that moral judgments merely express feelings and do not express any proposition that could be true or false at all. Others have noted that he does not actually declare that such deductions are impossible, but only that "vulgar systems of philosophy" are undermined by their failure to explain them, leaving open the possibility that the moral sense theory he is about to propound will explain them.

MORALITY AND VIRTUE

Excerpted from *Hume*

There is no need to interpret the famous passage by appeal to a supposed rejection of all moral truth. Hume writes frequently of "moral judgments," and these will typically be true or false: the idea of a moral trait as falling within the revival set of VIRTUE or VICE either will correspond to reality in consequence of that trait being properly represented within the idealized revival set of the concept in question or not [2.4, 5.2]. What he calls "moral obligation" to perform an action arises "when the neglect or non-performance of it displeases us" by means of sentiments of moral disapprobation (THN 3.2.5.4/517). Accordingly, he need only be claiming, in accordance with the thesis of the section as a whole, that mere inferences from matters of fact are not sufficient to render moral obligation explicable without also invoking, in addition, a further source of moral impressions, soon to be identified as the moral sense. His emphasis on the seeming divide between is and *ought* may also indicate, however, that the normative status of a concept [5.1] provides it, through its practical conceptual role and accompanying social and personal commitments, with an additional element of expressive meaning in addition to the element of meaning that results simply from having an idealized revival set.

Whereas *Treatise* Book 3 begins by answering the question of the roles of reason and sentiment in morals, the second *Enquiry* raises the question at the outset only to defer its answer to the work's first appendix. This is in keeping with the simplified and more directly practical character of the *Enquiries* generally, and Hume's presentation of the topic in the later work is also in keeping with his willingness in the *Enquiries* to engage in "reconciling" projects on at least some disputed issues [1.4, 6.4]. Thus, in Appendix 1 ("Concerning Moral Sentiment") of the second *Enquiry*, he begins by emphasizing the importance of reason in determining what the typical consequences of mental characteristics actually are. His thesis about the inability of reason alone to make moral distinctions remains unchanged, however, and he proceeds to offer five arguments for it. The first consists in a review of the unsuitability of both demonstrative and probable reasoning, a review similar to the concluding argument of the *Treatise* on the topic. The second is that one must first establish all of the particular causal circumstances of a case before making a moral determination by appeal to feeling. The third is that morality is analogous to aesthetics, where it is generally allowed that feeling, rather than reason alone, is the ultimate source of the distinctions made. The fourth is that inanimate objects can stand in the same relations discovered by reason as human beings can, but such objects cannot themselves be

MORALITY AND VIRTUE

vicious or virtuous. The final argument, echoing the Motivation Argument of the *Treatise*, is that reason alone cannot set ends, but only discover means to them, whereas virtue is preferred in itself.

As this last contrast implies, the success of the Motivation Argument requires that morality be directly motivating. Hume intends his account of the moral sense to explain how this can be so: moral distinctions originate in a sense the activation of which results in responses that are—unlike the responses of the color sense, for example—themselves pleasures or pains. For this reason, morality plugs directly into the faculty of the passions [3.6] in such a way as to produce desire, love, hatred, pride, humility, benevolence, anger— and thereby volition. In the absence of the resources provided by the faculty of the passions with which the moral sense interacts, reason as a separate inferential faculty could not engage the will by appealing to any object of "concern" [3.7] to the agent. For Hume, the fundamental problem with most vicious people is not that they are irrational or reason badly, but that their characters are marked by (other) vices that make them harmful and disagreeable.

8.2 Natural virtues

Just as Hume offers both a productive and a responsive definition of 'cause'—in terms of constant conjunction and inference-and-association, respectively [4.3, 6.3]—so too he offers both a productive and a responsive definition of 'virtue': "every quality of the mind, which is *useful* or *agreeable* to the *person himself* or to *others*" (EPM 9.12/277) and "*whatever mental action or quality gives to a spectator the pleasing sentiment of approbation*" (EPM App 1.10/289), respectively [4.2]. The productive definition provides a rough but ready fourfold classification of virtues: those useful to the possessor, those useful to others, those immediately agreeable to the possessor, and those immediately agreeable to others. Many virtues, of course, satisfy more than one of these descriptions. Indeed all virtues elicit love when present in others, on Hume's view, and love naturally elicits a desire to benefit the person loved; hence we may infer than any virtue is likely to prove useful to its possessor to at least some extent. Nevertheless, it is often clear that the approbation-eliciting power of a given trait is derived at least in the first instance either exclusively or predominantly from just one of these four sources.

Nearly all of the artificial virtues, for example, derive most of their original approbation-eliciting power from their usefulness to others. But such important natural virtues as benevolence and kindness also derive their approbation largely,

MORALITY AND VIRTUE

if by no means exclusively, from this source. Virtues that are predominantly useful to the possessor, in contrast, include: "prudence"; "discretion" (the capacity to give due attention to characters, circumstances, and available means in the conduct of life); "industry" (industriousness); "frugality" (as a mean between avarice and prodigality); "strength of mind" (a predominance in motivational force of calm passions above violent ones); "courage"; "wisdom and good sense" (the capacity to reason well, proportioning belief to evidence without falling victim to "unphilosophical" influences on belief); and even (strength of) "memory" (EPM 6). That human beings do morally approve these latter traits wherever they occur shows all the more clearly, in Hume's view, that morality is not based simply in self-interest. For in these cases, at least, the observer is not, unless considering his or her own character, among the primary beneficiaries of the trait; indeed, where the observer is in competition with the possessor, the traits may even be harmful to the observer and produce negative feelings from the separate and non-moral point of view of self-interest.

Virtues that are predominantly immediately agreeable to their possessors include: "greatness of mind" or "dignity of character"; "cheerfulness"; "tranquility"; and "delicacy of taste" (sensitivity to beauty and deformity in both discrimination and degree of sentiment) although this latter trait can also be painful at times. Courage, while predominantly useful to its possessor, is also agreeable to its possessor insofar as it restrains unpleasant fears. Benevolence, too, while predominantly useful to others, is warmly agreeable to its possessor; indeed, it is because of this agreeableness, Hume claims, that we tend to praise benevolence even when it somewhat exceeds the bounds of its usefulness (EPM 7).

Among virtues for which immediate agreeableness to others predominates, he mentions being a good and easy conversationalist (see also THN 3.3.4.9/611, with its distinctively Humean parenthetical addition that the personal merit derived in this way may "be very considerable"); "wit and ingenuity" (in a sense at least including the ability to produce humor [5.1]); "eloquence"; "modesty" (in the sense of a lack of arrogance and a disposition to give flattering attention to the opinions of others), especially in the young; "decency" (as "proper regard to the age, sex, character, and station" of others); and "cleanliness." He also mentions two virtues better known in the present day under shorter names. One is

> something mysterious and inexplicable, which conveys an immediate satisfaction to the spectator, but how, or why, or for what reason, he cannot pretend to determine. There is a MANNER, a grace, an ease, a genteelness,

MORALITY AND VIRTUE

> an I-know-not-what, which some men possess above others, which is very different from external beauty and comeliness, and which, however, catches our affection almost as suddenly and powerfully.
>
> (EPM 8.14/267; see also THN 3.3.4.11/612)

This, I take it, is what we would call *charisma*. He also mentions as immediately agreeable the talent for a "certain easy and disengag'd behaviour" (THN 3.3.1.27/589)—what we might call being *cool*.

Hume singles out as notable two different clusters of virtues: those relating to "greatness of mind" and those relating to "goodness or benevolence" (THN 3.3.2–3). The former include "courage," "intrepidity," "ambition," "love of glory," and "magnanimity." These "and all other shining virtues of that kind" derive their virtue primarily from their usefulness and/or agreeableness to their possessors and are further distinguished by their "strong mixture of self-esteem" or pride. "Nothing," he asserts, "can be more laudable, than to have a value for ourselves, where we really have qualities that are valuable" (THN 3.3.2.8/596). There are two primary reasons why such self-valuing traits elicit approbation. First, a due degree of pride gives one confidence in one's undertakings and thereby makes success in them more likely, while also inspiring the bold and enterprising projects on which fortune often smiles. Second, few passions are as immediately agreeable to the possessor as pride.

Due limits must be placed on the pride that is based on false self-assessments, however, and especially on the nature of its expression, because an "overweening" pride is immediately disagreeable to others. This disagreeableness, Hume explains, results from the combination of two mental operations. First, by the operation of sympathy, the observer involuntarily acquires from the proud person a feeling of that person's merit; then, by the operation of comparison [3.6] with the observer's opinion of himself or herself, the observer feels an unpleasant humility. In consequence, it is necessary to establish rules of politeness and "good-breeding" to avoid giving offense. Yet he remarks,

> I believe no one, who has any practice of the world, and can penetrate into the inward sentiments of men, will assert, that the humility, which good-breeding and decency require of us, goes beyond the outside, or that a thorough sincerity in this particular is esteem'd a real part of our duty. On the contrary, we may observe, that a genuine and hearty pride, or self-esteem, if well conceal'd and well founded, is essential to the character

MORALITY AND VIRTUE

> of a man of honour, and that there is no quality of the mind, which is more indispensibly requisite to procure the esteem and approbation of mankind.
>
> (THN 3.3.2.11/598)

Christianity's tendency to treat humility as a virtue and pride as a pagan vice is in his view positively pernicious.

Among the pleasurable indirect passions, pride has self as object, whereas the object of love is another person [3.6]. In contrast to the virtues of greatness of mind, which are closely related to pride, the virtues of goodness and benevolence are closely related to love. These latter virtues include "benevolence" itself (which, as a virtue, is a disposition toward the passion of benevolence or desire for the wellbeing of others), "generosity," "humanity," "compassion," "gratitude," "friendship," "loyality," and "liberality." They derive their merit primarily from their usefulness and agreeableness to others. Because the passion of love is itself immediately pleasurable, however, we also praise "whatever partakes of it" through an immediate sympathy with its possessor. Because we expect love to be partial to some extent—that is, to be felt more towards family and friends than to strangers—we accordingly expect and allow some partiality in the expression of these virtues as well. Hume holds that the virtues of greatness are properly regulated by virtues of goodness. While military glory, for example, inspires approbation when we sympathize at a distance with the proud military leader, benevolence for those whose lives are devastated by wars reduces our approbation; and courage and ambition unregulated by benevolence render their possessor fit only to be "a tyrant and a public robber" (THN 3.3.3.3/604).

Hume recognizes that his contemporaries often distinguish between "moral virtues" and "natural abilities," with the latter—such as "good sense," "quickness of apprehension," "strong memory," and charisma—considered merely as mental endowments without distinctive moral worth. He argues, however, that this distinction between kinds of personal merit is largely overrated (THN 3.3.4, "Of natural abilities"). The traits that are always considered moral virtues and the traits that are often classified as mere natural abilities can equally produce love and pride, he observes; we care equally about our reputations for both; and both cause benevolence and goodwill to be directed at their possessors. Although he concedes that the felt character of the particular sentiments of approbation they elicit may differ somewhat in many of these cases, the same holds true, he emphasizes, for different traits that are universally regarded as virtues, especially when these traits involve different mixtures of, or relations to, love and pride.

MORALITY AND VIRTUE

Excerpted from *Hume*

Because "legislators, and divines, and moralists" are concerned to improve the behavior of others through exhortations and through the additional motives of rewards and punishments, they often seek to draw and emphasize the distinction between moral virtues and natural abilities on the grounds that natural abilities and their exercises are not voluntary. Yet many of the recognized virtues of the great, Hume argues—such as fortitude, constancy, and magnanimity— are equally involuntary in their possession and largely in their expression as well. Furthermore, the more passionate a character is, the less under voluntary control its vices are—yet the vices are often all the more blamable for that.

Because the supposed line between virtues and natural abilities is vague and admits of no standard by which a clear distinction can be made, Hume declares the question of which traits belong to which category to be "merely verbal" (EHU Appendix 4, "Of Some Verbal Disputes"). In contrast, strength and beauty can be clearly distinguished from virtue as being physical, rather than mental, characteristics. He allows charisma as a virtue, we may assume, precisely because it is sharply distinguished from physical beauty and is discerned in a manner of conduct expressive of features (however "mysterious and inexplicable" they may be) of mind. Neglecting the distinction between virtues and natural abilities is another respect, he suggests, in which the ancient moral theorists are superior to the modern.

8.3 Justice as an artificial virtue

Hume most commonly uses the term 'justice' in a limited way that designates only respect for property. Although he sometimes adds the keeping of promises—which are also a person's "due"—to the scope of 'justice', he more commonly distinguishes promise keeping under the distinct term 'fidelity'. He also sometimes uses 'honesty' and even 'equity' in place of 'justice', seemingly just for verbal variation. For obedience to government, he generally uses the term 'allegiance'.

In Locke's political philosophy, which was especially popular with the Whig party of Hume's own time, there are divinely instituted moral obligations both to respect the property of others and to keep one's promises, including contracts, even in the absence of political society—that is, in "the state of nature"—and independent of any human conventions. The state of nature ends, on Locke's account, when a recognition of a need for more effective protection of property leads individuals to enter into a "social contract" to form a political society with one another and to establish a government from among the members of that

MORALITY AND VIRTUE

society. Their obligation to obey the government's edicts then results directly from their promise to one another to do so. Putting matters in Hume's terminology, then, Locke regards justice and fidelity as natural virtues, and the obligation to allegiance as derived from the obligation to fidelity. Hume, in contrast, declares that all three virtues are artificial, and that in no case is the obligation to one derived primarily from an obligation to another. On the contrary, he holds, they each derive their primary moral obligation directly from the same non-divine source: their usefulness to society.

In calling a virtue "artificial," it should be emphasized, Hume does not mean that it is any less a virtue or any less important than other virtues. Like all virtues, the artificial virtues are traits of character or mental characteristics that are useful or agreeable to their possessor or others, and which thereby elicit moral approbation from observers; several of the artificial virtues, moreover, are essential to human flourishing and survival. Rather, what makes them artificial is simply their dependence on the existence of a "convention." A convention exists among a group of individuals, as he explains it, when: (i) each individual has an interest in following a particular course of conduct, but only on the condition that the others follow a corresponding course of conduct; (ii) this common interest is mutually expressed and known among the individuals; and (iii) this mutual expression and knowledge serve to produce "a suitable resolution and behaviour" (THN 3.2.2.10/490). Individuals can enter into a convention without making a promise; on the contrary, promising is itself a specific convention. In Hume's elegant and well-known example, two people may pull the oars in a boat by convention without making any promises to each other, for each has an interest in pulling the oar on his side on condition that the other pulls the oar on the other side; each understands and expresses this common interest, verbally or non-verbally (perhaps simply by starting to row); and a suitable resolution and behavior results on the part of each (THN 3.2.2.10/490; EHU App 3.8/306–7).

Although it has often been misunderstood, Hume offers a positive and ingenious argument, sometimes called the Circle *Argument*, for his claim that justice is an artificial virtue (THN 3.2.1, "Justice, whether a natural or artificial virtue?"). Its starting point is a premise of his virtue ethics that we may call the *Virtue Ethics Thesis*:

> [Virtue Ethics Thesis:] [A]ll virtuous actions derive their merit only from virtuous motives, and are consider'd merely as signs of those motives.
>
> (THN 3.2.1.4/478)

MORALITY AND VIRTUE

'Motive' is used here in a broad sense that includes within its scope any mental traits that contribute to motivation.

Hume accepts this premise because—unlike many—he holds that VIRTUE, applied to traits of character, is the fundamental morally normative concept [5.1]; morally normative concepts can therefore apply to actions only in consequence of their being actions that persons with particular virtues or vices do or would perform. To this premise, he immediately adds what he regards as a simple point about explanatory priority: "An action must be virtuous before we can have a regard to its virtue." That is to say, human beings can sometimes be motivated to an action by the recognition of its morally normative merit, but in order for the act to have that merit, the action must already have derived the merit from the action's being the expression of some other morally approved trait or motive. He draws as a consequence from these two premises the *First Virtuous Motive Principle*:

> [First Virtuous Motive Principle:] [T]he first virtuous motive, which bestows a merit on any action, can never be a regard to the virtue of that action, but must be some other natural motive or principle.
>
> (THN 3.2.1.4/478)

The previous section of the *Treatise* ends with the observation that the term 'natural' may be opposed to 'rare', 'miraculous', 'civil', 'moral', or 'artificial', together with a remark that, in the remainder of the work, the context of each use of 'natural' will indicate the proper sense (THN 3.1.27–9n/473–76). In his statement of the First Virtuous Motive Principle, 'natural' clearly means "non-moral": a "natural motive or principle" is one that does not involve having a concern specifically for the moral merit of the action.

Hume's example of a "moral" motive, in contrast to a natural or non-moral one in this sense, is "duty," which he analyzes as the desire to perform actions of morally meritorious kinds, either (i) in the hope of acquiring the virtuous trait or motive that originally renders them morally meritorious, or (ii) in order to disguise from oneself one's lack of that virtuous trait or motive (THN 3.2.1.8/ 479). Because it is often useful to its possessor and others, duty is itself a virtuous motive, although always parasitic on the existence of other virtuous motives and often, for that reason, only a second-best one. It must be emphasized that a "non-moral" motive in this sense—that is, one that does not require a regard to the moral merit of actions—can still itself have moral merit as a virtuous mental trait. The strategy of Hume's argument is then to argue that, in the absence of a convention, there

MORALITY AND VIRTUE

would be no motive that could satisfy the First Virtuous Motive Principle with respect to just actions. For while some individual acts that are in accordance with the demands of property might be motivated either (i) by self-interest, (ii) by public benevolence (that is, concern for the interests or wellbeing of society or humankind), or (iii) private benevolence (concern for the interests or wellbeing of the individual most directly affected), none of these motives can explain what we might call the *full behavioral profile* of justice; each would often lead instead to actions that are contrary to the demands of property.

In order to discover the "first virtuous motive" to justice, then, we must understand how respect for property can be consistently motivated at all, and this requires understanding, in Hume's words, "how the rules of justice are established by the artifice of men"—that is, how property arises as a convention (THN 2.2.2, "Of the origin of justice and property"). Human needs are great, he observes, and the natural physical endowments of individual human beings are modest indeed in comparison with those of other animals. In order to survive and prosper, human beings must live in society, which allows them to augment their force by combining their strength, to augment their ability and skill by the division of their labor, and to augment their security through mutual aid. Happily, it is not necessary for human beings to foresee these advantages before entering into society, because the "appetite between the sexes" and "natural affection" for the resulting children are sufficient to institute society. Yet a serious threat to the maintenance of society lies in the combination of two features of human nature with two features of external circumstances. The two relevant features of human nature are "selfishness" (the tendency to prefer the satisfaction of one's own interests to those of others) and "limited generosity" (the tendency to prefer the satisfaction of the interests of one's family and friends to the satisfaction of the interests of others). The two relevant features of external circumstances are the scarcity of possessions (that is, items under a person's control) acquired by industry and the instability of possessions (that is, their liability to be taken by force to the advantage of the person taking them). In such circumstances, the industry of individuals will be naturally unavailing and individuals will commonly be threatened by violence.

The solution to the problem is a convention whereby individuals leave others in possession of the goods that they already possess on condition that the others do likewise for them. This is a coordinated course of conduct in which all parties benefit on condition that all participate; the common interest in this course of conduct can be mutually expressed and recognized; and a suitable resolution

MORALITY AND VIRTUE

and behavior result. While this convention is not "natural" in the sense opposed to "artificial," it is "natural" in the sense opposed to "rare," since its possibility and utility are readily discoverable by humans beings; as Hume remarks, any parent must establish it in order to maintain peace among his or her children (THN 3.2.2.14/ 493). Additional rules governing acquisition and the transfer of possession by consent are equally "natural" additions to the convention in this sense, also as a result of their obvious utility. Indeed, the human needs for rules of (i) stability of possession, (ii) transfer of possession by consent, and (iii) promise-keeping are so great, and the conventional rules themselves so certain to arise through human invention, that he allows them to be called "laws of nature" (THN 3.2.6.1/526)—although their moral force arises from the felt virtue of their observance, not any divine command. The details of such conventional rules may differ somewhat from one society to another, and somewhat arbitrary ways of rendering the rules precise will often be needed to avoid disputes. Effective rules, however, often appeal in some way or other to the principles of association of ideas that govern the imagination [2.2]. Through the adoption of rules of justice, what was mere pre-conventional "possession" becomes conventional "property."

The primary and original motive to the adoption of the convention of justice, Hume emphasizes, is self-interest, though supplemented by highly partial benevolence: it allows one to acquire and retain possessions for oneself, and also for the benefit of one's family and friends. In this way, he explains, self-interest comes to "restrain itself ": the very motive that threatened originally to destroy society through its unrestrained operation is recruited to redeploy itself in a more mutually beneficial direction. It does so by motivating the creation of a new convention and, at the same time, gives rise to a new motive that could not have existed before: the desire and standing disposition to govern or regulate one's behavior by the rules of property. To be sure, he notes, the full development and solidifying of this motive frequently depends on "repeated experience of the inconveniencies" (THN 3.2.2.10/490) of transgressing the convention. Nevertheless, the belief that such regulation is the most effective way to protect and increase one's possessions combines with previous desires for possessions to create a new desire to the perceived means. In coming to regulate one's behavior in this way, one undertakes to refrain from weighing up the specific advantages and disadvantages of following the rules of property before acting in each individual case, for such weighing would often lead to violations. Justice is a general policy or "scheme," one preferable relative to other schemes that might be considered and

MORALITY AND VIRTUE

preferable to having no scheme at all.

As the scheme first develops, human beings recognize only what Hume calls indifferently a "self-interested obligation" or a "natural obligation" ("natural" in the sense of "non-moral") to justice. As previously observed [8.1], he holds that one has a moral "obligation" to a course of action when one would feel moral reproach for oneself for failing to perform it (THN 3.2.5.4/517). Thus, the interested or natural obligation to justice presumably lies in the tendency to feel non-moral prudential reproach for oneself for harming one's own interests by not acting in accordance with the convention. Individuals soon come to reflect, however, on the beneficial effects of the system or scheme of justice on the wellbeing of members of society generally. As they do so, they feel sympathetic pleasure in considering the character trait of regulating one's conduct in accordance with the rules of justice, and this sympathetic pleasure, as may be expected, will in turn produce moral approbation for those who have that trait. In this way, human beings come also to recognize a distinctively moral obligation to justice, reproaching themselves through sentiments of moral disapprobation for not maintaining a course of conduct in accordance with its rules.

From the First Virtuous Motive Principle, Hume derives a corollary that he calls "an undoubted maxim": "*[N]o action can be virtuous, or morally good, unless there be in human nature some motive to produce it, distinct from the sense of its morality*" (THN 3.2.1.7/479). This maxim is often confused with the First Virtuous Motive Principle itself, but they make different demands and are satisfied by different motives. Self-interest is an original non-moral motive "in human nature" that becomes capable of producing the full behavioral profile of just actions through its own conventional self-restraint, and it thereby satisfies the Undoubted Maxim. Because it is not—given its other characteristic products— itself a virtuous motive, self-interest cannot itself satisfy the First Virtuous Motive Principle, but it also creates through convention a new motive: the desire and disposition to regulate one's conduct by the rules of justice. Although this new motive is not original in human nature, and so cannot satisfy the Undoubted Maxim, it is virtuous and so does satisfy the First Virtuous Motive Principle. Because justice depends for its existence on the convention that makes this motive possible, justice is an artificial virtue.

8.4 Other artificial virtues

Fidelity, in the sense of keeping one's promises, is a second artificial virtue,

MORALITY AND VIRTUE

dependent on a convention of its own (THN 3.2.5, "Of the obligation of promises"). In support of this thesis, Hume seeks to establish two distinct but related propositions: (1) "*that a promise woul'd not be intelligible before human conventions had established it,*" and *(2) "that even if it were intelligible [without a convention], it wou'd not be attended with any moral obligation*."

In defense of the first proposition, Hume offers two arguments. First, there is no "natural"—that is, in this context, non-artificial or non-conventional—act of the mind corresponding to the words, "I promise." It cannot be the formation of a resolution or intention, since this alone creates no new obligation. It cannot be desire, since we can promise to do what we do not desire to do; and it cannot be willing or volition, since that concerns the present rather than the future. Hence, the act of mind expressed by "I promise" can only be understood in terms of a convention. Second, promising is a way of voluntarily acquiring a new obligation to perform an action; yet obligation, as he has explained it [8.1, 8.3] depends essentially on sentiments, which cannot themselves simply be willed into existence. The creation of the new obligation, he proposes, can only be explained as the consequence of establishing a new relation between an action and an existing convention to which sentiments are already attached.

Hume offers two arguments in defense of the second proposition as well. The first argument invokes the same considerations as the previous argument: even if the mind could will a new obligation through promising, the obligation could have no moral character except through a relation to moral sentiments derived from the moral status of a convention. The second invokes the Undoubted Maxim that Hume has already derived from the First Virtuous Motive Principle [8.3]. That maxim requires that there be a motive in human nature to any virtuous act, yet in the absence of a convention there would be no motive capable of motivating the full behavioral profile of fidelity. The Undoubted Maxim can be satisfied only if the motive of self-interest, while not itself a virtuous one, can explain the creation of a convention that then generates a new motive capable of winning moral approbation and thereby conveying derivative moral approbation to the acts that manifest it. This is, in effect, an application of the Circle Argument to the case of promising.

Like the convention of property, the convention of promising arises, according to Hume, in response to a practical problem. In this case, parties who bear no particular affection for each other could each benefit from a mutual exchange of favors or possessions; but where this exchange cannot be

MORALITY AND VIRTUE

simultaneous, the exchange seemingly will not take place because the first party to perform has no security about the later performance of the second party. In Hume's elegant example, two farmers may have crops maturing at different times, the harvesting of which require the efforts of both parties. The solution is a convention employing a particular form of words or other expression: all agree to carry out any action described with that form of expression, on the understanding that failure to do so will debar the offender from future participation in the convention with all its benefits. As in the case of the convention of property, the convention of promising introduces the possibility of a new motive related to it, consisting in a desire and standing disposition to regulate one's action in accordance with the rules of the convention. Consideration of the effects of this character trait on others leads again, through sympathy, to moral approbation. The self-interested obligation to the keeping of promises is thus supplemented by a moral obligation, and fidelity is recognized as a virtue.

In the concluding section of the second *Enquiry*, Hume writes, "Having explained the moral *approbation* attending merit or virtue, there remains nothing, but briefly to consider our interested *obligation* to it" (EPM 9.14/278). He notes that everyone will naturally want to have traits useful or agreeable to themselves, and that vanity even by itself is sufficient to make one desire to have traits that are immediately agreeable to others. This leaves only the interested obligation to the traits useful to others. Of these, the more sociable, such as benevolence, are among the best and surest paths to happiness; we must care about something in order to derive any enjoyments from life, and caring about the wellbeing of others proves to be an excellent source of enjoyment. He then turns to justice and fidelity, writing in a famous passage:

> Treating vice with the greatest candour, and making it all possible concessions, we must acknowledge, that there is not, in any instance, the smallest pretext for giving it the preference above virtue, with a view to self-interest; except, perhaps, in the case of justice, where a man, taking things in a certain light, may often seem to be a loser by his integrity. And though it is allowed, that, without a regard to property, no society could subsist; yet, according to the imperfect way in which human affairs are conducted, a sensible knave, in particular incidents, may think, that an act of iniquity or infidelity will make a considerable addition to his fortune, without causing any considerable breach in the social union and confederacy. That honesty is the best policy, may be a good general rule;

MORALITY AND VIRTUE

> but is liable to many exceptions: And he, it may, perhaps, be thought, conducts himself with most wisdom, who observes the general rule, and takes advantage of all the exceptions.
>
> I must confess, that, if a man think, that this reasoning much requires an answer, it will be a little difficult to find any, which will to him appear satisfactory and convincing. If his heart rebel not against such pernicious maxims, if he feel no reluctance to the thoughts of villainy or baseness, he has indeed lost a considerable motive to virtue; and we may expect, that his practice will be answerable to his speculation. But in all ingenuous natures, the antipathy to treachery and roguery is too strong to be counterbalanced by any views of profit or pecuniary advantage. Inward peace of mind, consciousness of integrity, a satisfactory review of our own conduct; these are circumstances very requisite to happiness, and will be cherished and cultivated by every honest man, who feels the importance of them.
>
> (EPM 9.22–23/282–83)

In interpreting this passage, it is important to bear several points in mind. First, Hume is not proposing that moral obligation itself must be derived from or justified by self-interest. As he says, he has already explaining the moral approbation—and hence the moral obligation—attaching to virtue. Moral obligation applies regardless of whether there is also a self-interested obligation to virtue. Someone who behaved virtuously even at the expense of his or her own interests would not be behaving *irrationally* or even *unreasonably* in any sense by preferring virtue to self-interest, although such action might be *imprudent* [3.7]. Nevertheless, Hume holds that there is an interested obligation to virtue, even in the case of justice and fidelity, and that it is of great practical value to show that this is true. Were it not true, he remarks, someone who showed its falsity might be a good philosopher but could hardly be considered a friend to humankind.

Second, Hume is directly confronting an alternative policy to that of strict adherence to the rules of justice and fidelity—namely, the *knavish* policy of adherence to the rules whenever a violation would be detected or would carry no advantage in the particular case, while taking advantage of opportunities to violate the rules in other cases offering secrecy and advantage. Note, however, that even if this policy were in fact more beneficial than the policy of strict justice from the standpoint of self-interest, this need not undermine his account of the causal origin of the conventions of justice and fidelity in the motive of self-interest, so

MORALITY AND VIRTUE

long as the alternative policy of sensible knavery was not considered and judged to be more beneficial in the original circumstances in which the convention arose.

Third, Hume is conceding that someone who lacked developed sentiments of moral approbation for justice and fidelity would also lack a particular motive to uniform adherence to justice and fidelity that others possess. In the absence of such a motive, therefore, such a person's own self-interest night be better served by the knavish policy than by the policy of strict adherence, although this is by no means guaranteed: as Hume points out, the danger that a sensible knave will be detected despite his or her best efforts is ever-present, and the security of knowing that such an event cannot befall the follower of strict adherence is a further self-interested advantage to the latter.

Finally, however, Hume insists that one who does feel approbation for justice and fidelity will enjoy a distinctive "peace of mind" and a positive pleasurable pride in "making a satisfactory review of one's own conduct" that are simply unavailable to sensible knaves, who must be aware that others would disapprove their characters were they known. For this reason and others, it is better, even from the point of view of self-interest, to maintain the virtues of justice and fidelity.

The advantages provided by justice and fidelity—including the honoring of commercial contracts—allow societies to grow larger. Yet the larger the society, the easier it becomes to violate rules of property and promise-keeping without detection. The natural human tendency to prefer nearer but lesser goods to greater but more distant ones will then lead many to violate the rules on occasion, and the recognition of that tendency will lead members of society to lose confidence in the continued adherence of others, undermining the conventions themselves. Dissatisfaction with the results of this tendency, aided crucially by the ability to prefer the greater good at times when *both* are quite distant, facilitates a conventional solution: the setting up of individuals whose own self-interest will be to enforce the rules of justice and fidelity, and a general conventional agreement to obey those individuals (THN 3.2.7, "Of the origin of government"). Governments provide not only greater security in the execution of justice and fidelity, but also more equitable and impartial decisions in cases of dispute. Equally important, they serve to coordinate individuals in pursuit of further advantages such as large-scale public works, for which only a government can manage the coordination of the large numbers of workers required.

In the case of many early societies, Hume allows, the original choice of a

MORALITY AND VIRTUE

political leader may be accomplished through a promise, and for those who actually participate in it, such a promise can provide an additional moral obligation. A promise cannot bind later generations who are not parties to it, however, and the primary moral obligation to allegiance to government results from its direct usefulness to society; indeed, allegiance could be a virtue in the absence of any convention of promising at all. As with justice and fidelity, allegiance too improves human life by redirecting original passions through convention into more beneficial paths. Because allegiance is a virtue only through its utility, however, "an egregious tyranny in the rulers is sufficient to free the subjects from all ties of allegiance" by undermining its utility (THN 3.2.9.1/549).

Hume describes chastity (which he also calls "marital fidelity") and modesty (in the sense of sexual modesty, rather than the sense contrasted with proud deportment) as artificial virtues (THN 3.2.12, "Of chastity and modesty") as well. The institution of marriage, to which chastity refers, is clearly a convention; and he accepts the common suggestion that the primary purpose of marital fidelity is to provide security of paternity to fathers and thereby to support their affection for, and care of, children. The convention of modesty— manners of "backwardness" with respect to expressions, postures, and liberties relating to "the appetite of generation"—is intended to contribute to, and maintain the force of, chastity. While these are virtues for both sexes, he regards them as being of greater value in women because of the greater utility to society of their adherence. In this respect, he compares what he regards as the lesser obligation of men to chastity and modesty to the lesser obligation of princes to respect property and promises in their dealings with other princes (THN 3.2.11, "Of the laws of nations").

8.5 Moral diversity

Hume addresses diversity in moral judgment most directly in "A Dialogue," which appears, unnumbered, after the numbered appendixes to the second *Enquiry*. The character of Palamedes recounts the strange manners and morals encountered on a trip to the nation of "Fourli," which he soon reveals to be modeled on the ancient Greeks and Romans. Hume's first-person narrator responds that the same underlying principles of human nature are responsible for the differences in judgments, much as "the RHINE flows north, the RHONE south; yet both spring from the same mountain, and are also actuated, in their opposite directions, by the same principle of gravity" (EPM Dialogue 26/333). In the course of his writings, he discusses several different kinds of moral diversity.

MORALITY AND VIRTUE

Excerpted from *Hume*

One kind of moral diversity arises from differences of opinion concerning the causal consequences of characters or actions, matters of fact about which one party may be right and the other wrong. Since moral sentiments result chiefly from sympathy with those affected, according to Hume, such different opinions about the likely consequences of characters and actions can easily result in differences in moral judgment. Moral judgments about the willingness to assassinate tyrants, for example, differ in accordance with different factual judgments about the political consequences of both actual tyrannicide and the threat of its occurrence (EPM 2.19/180–81). The consequences of dueling provide a similar example (EPM Dialogue 34/335). Differing judgments about the moral permissibility of suicide often reflect different opinions about whether a deity has forbidden it and hence about the consequences of the act itself (EPM Dialogue 35/335). Just as the varying causal judgments at issue may be correct or incorrect, so too may the varying moral judgments depending on them.

In a second kind of moral diversity, one moral judgment may be correct and another incorrect through one party's greater capacity to approximate or anticipate the result of judging from the idealized point of view and with the idealized qualities of the standard of judgment [4.1]. Victors in war, for example, may sympathize with their leaders to the exclusion of more distant sufferers (THN 3.3.2.15/600–01), and other failures of sympathy may result from an exaggeration of superficial differences among persons that obscure their more fundamental human similarities (THN 2.1.11.8/319– 20). In cases of these kinds, it seems reasonable to suppose, there may be moral progress in the application of an existing standard of judgment. Hume further remarks at the conclusion of "A Dialogue" that "religious superstition or philosophical enthusiasm"—he mentions Blaise Pascal (1623–62) and Diogenes (4th–3rd centuries BCE), respectively—can lead to the adoption of principles of conduct that so remove a person from the ordinary circumstances of human life as to prevent the operation of the "natural principles of the mind." In such cases, the sensitive qualities of individuals are diminished or perverted, hindering them from making sound judgments in much the same way that a visual deficiency undermines color judgments. It may also be more difficult for others to judge such products of "artificial lives" accurately, precisely because the removal from ordinary circumstances of human life makes it more difficult to trace the consequences of their unusual characters.

In a third kind of case, by contrast, the apparent disagreement may be resolved through relativization, allowing that the judgments of both parties are

MORALITY AND VIRTUE

positively correct about the application of morality to their own social or cultural circumstances when differences in them give traits differing degrees of typical usefulness. (As remarked previously, Hume seems to allow a comparable relativization of probability judgments to different experiential bases [4.4]). Military courage, for example, is of greater utility, and is therefore a greater virtue, in societies in which individuals are constantly exposed to the threat of invasion and destruction from outside forces (EPM Dialogue 39/337). Many differences in conventional customs—for example, concerning modesty or forms of government—may themselves be morally innocent, but render particular character traits more or less useful in consequence (EPM Dialogues 49–51/340–41). Similarly, it is important that there be conventions governing such matters as the conveyance of property and the degree of blood relationship allowed in marriage, but it is often arbitrary exactly where the lines are to be drawn; in such cases as these, the virtuous disposition is the one that regulates behavior in accordance with the convention actually in place (THN 3.2.3, "Of the rules, which determine property"). In extreme circumstances, such as a shipwreck or severe famine, the utility of any convention of justice may itself fail, bringing with it a release from the moral obligation to regulate actions in accordance with such a convention and making the regulating trait itself no longer a virtue (THN 3.2.2.16/494; EPM 3.12/188).

In a final kind of case, however, diversity in judgment is blameless in consequence of lack of precision in the standard of judgment itself [4.2]. Most notably, Hume urges, there are sometimes differences in response from sympathetically weighing degrees of tradeoffs between competing goods and harms that are not resolved by the standard. The relative values of greatness of mind as contrasted with goodness and benevolence may be one such case. Another, emphasized in "A Dialogue," lies in the different degrees of value that, he reports, the French and the English place on marital fidelity and "the gallantry of amours and attachments." Whereas the French sacrifice some of the utility of domestic fidelity and constancy in order to experience more of the agreeable qualities associated with ease, freedom, and openness of social commerce, the English sacrifice some of the agreeable for the sake of greater utility (EPM Dialogue 47–48/339–40).

Blameless diversity in balancing the degrees of virtues that may sometimes be in tension leads readily to blameless moral diversity in action. For example, while the rules of justice may themselves be exceptionless, and the virtue of

MORALITY AND VIRTUE

justice consists in the disposition to regulate one's conduct by those rules, the virtue of justice does not necessarily trump all other virtues in value (THN 3.2.11.3/568). Under extreme circumstances—say, an act of returning property that would result in the destruction of all life on the planet—the virtuous person, we may expect, would display benevolence rather than strict adherence to the rules of justice. In more difficult cases, however, it might well be that either of two contrary actions might be at least permissible: one as displaying an admirable benevolence despite its troubling injustice, and the other as displaying an admirable commitment to justice even at the worrying expense of ignoring the call of benevolence.

Hume's moral theory allows for moral progress in several different respects. First, of course, there may be purely practical progress: recognized virtues may become more prevalent and vices less so, perhaps as the result of improved conditions of sociality and perhaps even as the result of the exhortations of moralists or the argued defenses of philosophers such as Hume. Second, there may be improvements in the science of morals itself. Moral distinctions may be more accurately drawn through increased understanding of human life and psychology, or through improved capacity to sympathize and otherwise imaginatively take up the point of view of the standard of morals. The proper scopes of relativization and blameless diversity may also come to be better understood. All such cognitive improvements may also lead to improvements in character. Finally, and more radically, the normative concepts VIRTUE and VICE themselves may continue to develop through time as new qualities or features of points of view are incorporated through convergence into the standard of judgment. This is comparable in some ways to the ongoing development of the concept PROBABILITY through the adoption of new rules of probability and the refinement of its standard in the centuries since Hume. In the case of morals, at least, changes in the standard of judgment may even be reflexively responsive to the felt moral value of adopting and honoring such refinements of the standard—or so we may hope.

8.6 Conclusion

Few philosophers have had more influence on recent and contemporary moral philosophy than Hume—both for what he actually argued and maintained and for what he has been thought to have argued and maintained. Hume's Law, interpreted as a formulation of a strong distinction between facts and values, is one example.

MORALITY AND VIRTUE

Excerpted from *Hume*

Another lies in his role in debates about moral truth and moral motivation. Explicitly appealing to Hume's inspiration, Michael Smith has provided an influential formulation of what he calls "The Moral Problem," consisting in the incompatibility of three individually attractive theses:

> *Moral Cognitivism*: Moral judgments express beliefs that are true or false.
>
> *Moral Internalism*: Moral judgments are intrinsically motivating (to at least some extent).
>
> *Belief/Desire Motivation*: No belief is intrinsically motivating; motivation always requires a corresponding desire or passion.

It is sometimes suggested that Hume's solution to the problem would be to deny Moral Cognitivism [8.1]. Yet while moral sentiments— impressions of moral approbation or disapprobation—are not themselves true or false, the predicative moral judgments that employ moral concepts in accordance with a standard of judgment clearly are. Hume's own solution to the problem, I suggest, would be to distinguish the respects in which moral judgments are, and are not, "intrinsically" motivating.

First, it is not a metaphysically necessary truth (that is, not a relation of ideas in Hume's terminology) that sense-based judgments of character traits or actions motivate, at least if these judgments are understood narrowly as the classification of a character or action in the revival set of an abstract idea [2.4]. No substantive questions about actual causal relations can be *a priori* for Hume, and the ability of such judgments of characters and actions to contribute to the causation of volitions can only be a matter of fact dependent on human psychology and discovered by experience. Furthermore, given appropriate information or evidence, it is causally possible to make a particular predicative moral judgment, employing moral concepts such as VIRTUE and VICE, without at that very moment feeling any motivating force; in a similar way, it is possible, given appropriate information, to make a particular conceptualized predicative color judgment about an object without at that moment actually seeing it or receiving any sensation from it. On the other hand, the moral sentiments that are the ultimate source of moral distinctions and judgments are themselves pleasures and pains. In normal human psychology, these sentiments are therefore of precisely the right kind to exert at least some causal influence on volition whenever they occur. Furthermore, it is part of the attributed practical relations pertaining to the conceptual role of normative concepts generally that terms expressing them are

MORALITY AND VIRTUE

"taken in a good [or bad] sense" [5.1], and anyone who uses such terms without some appreciation of and commitment to the qualities falling under the concepts has not fully incorporated their normative character into her or his own usage of the terms. When "intrinsic motivation" is understood in this limited psychological way, Moral Internalism is fully compatible with both Moral Cognitivism and Belief/Desire Motivation.

Hume's distinctive combination of moral sense theory and virtue ethics helps to justify other aspects of his moral theory as well. Partly because it is sense-based, it is *broad*, encompassing all kinds of mental traits, self-affecting as well as other-affecting, intellectual as well as affective, involuntary as well as voluntary. It is not *deontological,* or duty-centered: while Hume recognizes a motive of duty, and even allows it to be a kind of virtue, it is generally a second-best motive from a Humean moral perspective. The best parent, for example, acts from natural affection for children, the "first virtuous motive" to childcare; the parent who cares for his or her children only from duty is typically seeking either to hide the absence of this motive from himself or herself, or seeking to inculcate or encourage it through practice. It is also not *consequentialist*, or consequence-centered, however: while the consequences of actions typical of a mental trait play a crucial role in determining whether the trait will be approved by the moral sense, it is the moral sense itself, not the balance of consequences, that is the source of moral evaluations, and that sense fundamentally evaluates characters, not consequences.

Because Hume's moral sense is primarily responsive to pleasure and pain, and especially to sympathetic pleasure and pain, his moral philosophy is *non-austere*, valuing traits that are conducive to pleasure and disvaluing those—such as the "monkish virtues" of "celibacy, fasting, penance, mortification, self-denial, humility, silence, and solitude" [1.4, 9.7]—that are conducive to pain. On the contrary, in a flight of personification he writes:

> But what philosophical truths can be more advantageous to society, than those here delivered, which represent virtue in all her genuine and most engaging charms, and make us approach her with ease, familiarity, and affection? The dismal dress falls off, with which many divines, and some philosophers have covered her; and nothing appears but gentleness, humanity, beneficence, affability; nay even, at proper intervals, play, frolic, and gaiety … . And if any austere pretenders approach her, enemies to joy and pleasure, she either rejects them as hypocrites and deceivers; or if she

MORALITY AND VIRTUE

> admit them in her train, they are ranked, however, among the least favoured of her votaries.
>
> (EPM 9.15/279)

For similar reasons, Hume's moral theory is also *non-rigorous*: while artificial virtues require regulating one's behavior by conventional rules without weighing the advantages and disadvantages in individual cases, the virtue of adhering depends on the utility of the convention and does not extend beyond the circumstances in which the convention retains its general utility. In his treatment of pride (and elsewhere) he treats the obligation to truthfulness as properly limited by the demands of sociability and politeness. Moreover, his moral theory is pluralist: the sense-based character of his moral philosophy, drawing as it does on an idealized point of view and respondent qualities, readily allows for some blameless diversity in judgments of character, and its treatment of character traits as the primary objects of evaluation for the fundamental normative concepts allows for further latitude of blameless diversity in the moral judgment of actions, which are evaluated only as signs or expressions of character. Finally, as we shall have greater occasion to observe in the next chapter, his moral theory is entirely *secular*: at no point does he appeal to any religious doctrines in the making of moral judgments.

In its independence of a deity, Hume's moral philosophy is naturalistic, and it is also naturalistic inasmuch as it does not invoke any explanatorily basic normative qualities; the normativity attaching to moral concepts is explicable instead by appeal to their actual role in human life. It is empiricist as well, inasmuch as it draws moral concepts and judgments from experience, without postulating substantive a priori moral principles [5.4]. Although they also typically serve—like all normative judgments—to express a social and personal commitment, Humean moral judgments can classify correctly or incorrectly, and hence they can be true or false. As such, they presumably fall within the scope of mitigated skepticism's universal diminution of degrees of probability. His moral theory is not radically skeptical, however. Instead, just as he rejects radical skepticism about the reality of aesthetic distinctions (such as that between the literary genius of "Ogilby and Milton" in "Of the Standard of Taste," EMPL I.23: 230), and ultimately rejects skepticism about the reality of distinctions of probability [7.3], so too he rejects radical skepticism about the reality of moral distinctions. Indeed, just as overcoming radical skepticism about the reality of aesthetic distinctions and probability distinctions requires the passage of time and the

MORALITY AND VIRTUE

return of sensibility, so too does the overcoming of radical skepticism about moral distinctions:

> Those who have denied the reality of moral distinctions, may be ranked among the disingenuous disputants; nor is it conceivable, that any human creature could ever seriously believe, that all characters and actions were alike entitled to the affection and regard of every one. The difference, which nature has placed between one man and another, is so wide, and this difference is still so much farther widened, by education, example, and habit, that, where the opposite extremes come at once under our apprehension, there is no scepticism so scrupulous, and scarce any assurance so determined, as absolutely to deny all distinctions between them, Let a man's insensibility be ever so great, he must often be touched with the images of RIGHT and WRONG; and let his prejudices be ever so obstinate, he must observe, that others are susceptible of like impressions. The only way, therefore, of converting an antagonist of this kind, is to leave him to himself. For, finding that no body keeps up the controversy with him, it is probable he will, at least, of himself, from mere weariness, come over to the side of common sense and reason.
>
> (EPM 1.2/169–70)

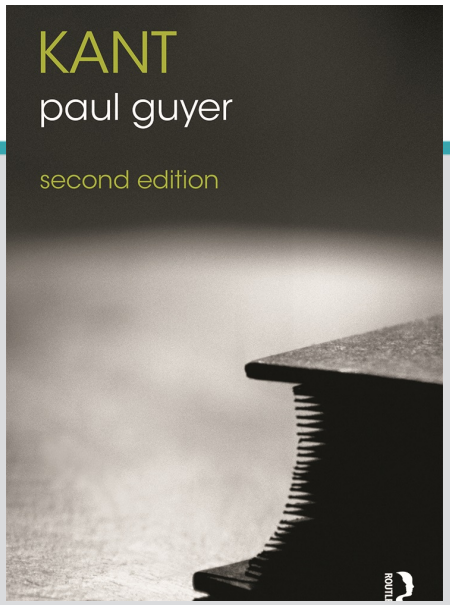
The final sentence of the *Treatise* pronounces moral precepts to constitute a "science" [5.4], confirming the initial ambition of its Introduction to contribute to the science of "Morals" through its development of the science of man [1.2].

CHAPTER

4

LAWS OF FREEDOM

THE FOUNDATIONS OF KANT'S MORAL PHILOSOPHY



LAWS OF FREEDOM

Excerpted from *Kant*

We now turn from the abstraction of Kant's philosophy of science to his practical philosophy, which can seem equally remote from our everyday experience. Kant is famous for the derivation of an apparently formalistic fundamental moral law from the most abstract and austere premises. He begins his *Groundwork for the Metaphysics of Morals* (1785) with the claim that the only thing of unconditional value is a good will, argues that such a will manifests itself only in doing one's duty for its own sake, and then concludes that since doing duty for its own sake deprives the will of any object of desire as a reason for action, nothing is left as a possible principle of morality "but the conformity of actions as such with universal law" (G, 4:402). In the second section of the same work, he maintains that "moral laws are to hold for every rational being as such" and must therefore be derivable from the very "universal concept of a rational being as such" (4:412). In the *Critique of Practical Reason* (1788), he premisses that a moral law must be completely necessary and universal and then concludes that only a moral principle that is entirely formal and makes no reference to any object of desire can satisfy that requirement. Specifically, he argues that genuine moral laws or "practical principles"must hold "for the will of every rational being as such" (CPracR, 5:19), that all "practical principles that presuppose an object (matter) of the faculty of desire as the determining ground of the will are, without exception, empirical and can furnish no practical laws" (5:21), and thus that "If a rational being is to think of his maxims as practical universal laws, he can think of them only as principles that contain the determining ground of the will not by their matter but only by their form" (5:27).We will give these arguments a hearing shortly, but it seems clear from the outset that they presuppose what might be a controversial assumption about what a moral law must be like, and it is by no means obvious how they could be expected to gain a grip on the moral sensibilities of ordinary human beings.

Elsewhere, however, Kant suggested a more intuitive foundation for his moral philosophy. In the classroom lectures on ethics that he gave during the decade before he began publishing the works just mentioned, he is reported to have argued that "Freedom . . . is the capacity which confers unlimited usefulness on all the others" and therefore is "the highest degree of life," the "inner worth of the world," but that "insofar as it is not restrained under certain rules of conditioned employment, it is the most terrible thing there could be"; in order to realize its potential value, therefore, freedom must be exercised in accordance with a rule "under which alone the greatest use of freedom is possible, and under which it can be self-consistent" (LEC, 27:344, 347). The rule is simply that freedom

LAWS OF FREEDOM

must be "consistent with itself," that is, that my use of freedom on one occasion be consistent with my continued use of it on all other possible occasions, and that my use of freedom be consistent with everyone else's use of their freedom. Of course, to state this rule at such a level of abstraction is easy; to say what it actually requires of us in the concrete circumstances of human life considerable thought will be required. That is why we must employ our reason to formulate the moral law in a variety of forms and then to derive a detailed system of duties from them. But on this approach, we do not have to begin with the completely abstract idea that rationality as such is of intrinsic value or that there is some inexplicable necessity for acting in accordance with a necessary and universal law. Instead, as Kant put it in lectures on "natural right" (political philosophy) that he gave during the very semester when he was composing the *Groundwork*, "If only rational beings can be an end in themselves, this is not because they have reason, but because they have freedom. Reason is merely a means" (NFey, 27:1321). That is, through reason we grasp the rules that we need to follow in order fully to realize our freedom as autonomy, or "the property that a will has of being a law to itself" (G, 4:447).

Of course, one might well think that the claim that freedom itself is our most fundamental value could use some support. Kant sometimes wrote as if this is an obvious truth about human psychology. In some notes that he made in his own copy of his early work*Observations on the Feeling of the Beautiful and Sublime*, Kant wrote:

> The human being has his own inclinations, and by means of his capacity of choice a clue from nature to conduct his actions in accordance with these. Nothing can be more appalling than that the action of a human stand under the will of another. Hence no abhorrence can be more natural than that which a person has against servitude. On this account a child cries and becomes bitter if it has to do what another wants without having made an effort to make that pleasing to him. And it wishes only to become a man quickly and to operate in accordance with its own will.
>
> (NF, pp. 10–11)

This makes it sound as if the love of freedom is a basic trait of human psychology, and thus that the moral force of laws for the realization of freedom ultimately comes from a fact about human nature. It is not clear that such a foundation for morality would be consistent with Kant's insistence that the moral law must be

LAWS OF FREEDOM

valid for every rational being, human or otherwise, thus that "a pure moral philosophy" must be "completely cleansed of everything that may be only empirical and that belongs to anthropology" (G, 4:389). But it is also not clear whether Kant really has an alternative but equally gripping account of the normative force of the moral law, so this psychological assumption may play an indispensable role in Kant's subsequent moral philosophy even if he does not acknowledge it.

In what follows, our first order of business will be to examine the arguments that Kant made for his formulation of the moral law in his mature published works, then to see how his earlier idea of the inner worth of freedom reappears in his mature works and how the various formulations of the fundamental principle of morality that he offers in those works can be understood as formulations of the rules necessary in order to realize the value of freedom.As autonomy in its practical sense is nothing other than freedom achieved and sustained through its adherence to law, this will constitute the next step in our study of Kant's overarching conception of autonomy. Then we can return to the question of how or even whether Kant can argue for his fundamental normative assumption or conception of value.

The derivation of the categorical imperative

The fundamental principle of morality, Kant has claimed, must be unconditionally valid for any rational being. If any being were perfectly rational, it would automatically act in accordance with this law, and the law would therefore not appear to be a constraint. But we human beings are not perfectly rational, and thus although we recognize the unconditional validity of the moral law, it also appears as a constraint to us, something that may be in conflict with our irrational side. The fundamental principle of morality thus presents itself to us in the form of a "categorical imperative": categorical, because we recognize that its demands are unconditional, but an imperative, because we recognize this law as something we *ought* to follow, thus as a constraint, that is, not something we always *want* to follow.The concept of the categorical imperative is thus not identical to the concept of the fundamental principle of morality, but is rather the way in which the fundamental principle of morality presents itself to us as beings who are rational but not purely rational. But Kant takes it to be obvious and not in need of any special argument that we will often experience the stringent demands of morality as a constraint; thus, although his arguments are aimed at a derivation of the categorical imperative, all of his effort is aimed at demonstrating the content of

LAWS OF FREEDOM

the fundamental principle of morality and proving that it is valid or binding for us, not at reminding us that we often experience that validity as a constraint.

Kant discusses the derivation of the categorical imperative at length in the *Groundwork*, and then more briefly in the *Critique of Practical Reason*, which is devoted primarily to the problem of free will and then, under the topic of what Kant calls the "highest good," to reestablishing a relation between virtue and happiness that he seems to have severed completely in the *Groundwork*. The *Groundwork* is divided into three sections, which Kant labels respectively the "Transition from common rational to philosophical moral cognition" (G, 4:393), the "Transition from popular moral philosophy to the metaphysics of morals" (4:406), and the "Transition from metaphysics of morals to the critique of pure practical reason" (4:446). He does not mean the same thing by "transition" in each case: while the argument of the second section is that "popular moral philosophy" must be *replaced* by a philosophically sound "metaphysics of morals," the first and third sections argue that this metaphysics of morals must be *grounded* in *both* genuine common sense and a philosophically sophisticated "critique of pure practical reason." However, this organization of his arguments is also in some tension with another claim that Kant makes, namely that in the first two sections he is just *analyzing* the content of the fundamental principle of morality for any rational beings, and that it is only in the third section that he will show that this principle *applies* to us as the categorical imperative (see 4:392, 425). The tension is that Kant at least tacitly supposes that sound common sense always knows both what the categorical imperative requires and that it requires that of us, not needing a subtle philosophical argument to prove that. In the *Critique of Practical Reason* Kant will resolve this tension in favor of common sense when he asserts that our consciousness of our obligation under the moral law is a "fact of reason" from which the freedom of our will may be inferred but which cannot itself be deduced from any more fundamental premise (CPracR, 5:29–31).We will return to this issue, but for now let us follow the opening arguments of the *Groundwork*.

Kant begins his analysis of "common rational moral cognition" by arguing that common sense recognizes that the only thing of unconditional value is a good will. He argues first that gifts of nature and fortune, such as strength, talent, and resources, are not unconditionally valuable, because whether they are good or evil depends on whether they are put to use by a good or evil will (G, 4:493–4). This is indeed a bit of common sense, but it does not imply, as Kant seems to think, that a good will is of any value by itself, entirely independently of "what it effects or

LAWS OF FREEDOM

accomplishes" (4:394). More importantly, it does not tell us anything about the content of the good will or the principle by which it is governed beyond the obvious fact that a good will cannot simply be the will to possess goods of nature or fortune. Kant's next argument, that the point of a good will cannot be to produce *happiness* because it is not particularly good at doing that (4:395), is more important, but it rests on the teleological premise that each of our faculties is naturally intended for one purpose and that it must be good at that purpose; as we saw in Chapter 4, Kant relies on this principle in his general theory of the function of reason, but it could certainly be questioned. (Kant will provide a better account of why the principle of morality cannot simply be to seek (or maximize) happiness in the *Critique of Practical Reason*.)

Having made these opening sallies, Kant then offers a more careful analysis of the common conception of what it is to have a good will. He argues first that a person demonstrates possession of a good will not just by performing an action that is *in conformity with duty,* but by performing such an action *from duty* (4:397–8). In other words, a person with good will does not just do what duty requires but is also motivated by the recognition that the action is her duty or by the general principle to perform an action if and only if it is her duty. We are supposed to recognize this from such common examples as the honest shopkeeper: if a shopkeeper refrains from cheating even his most inexperienced customers because he thinks that a reputation for honesty will be good for his business in the long run, that is just action out of self-interest, for which to be sure he cannot be criticized, but for which he also does not earn our esteem, because he does not demonstrate a good will (4:397). Second, Kant claims that it follows from this that the moral value of an action cannot lie in the end or state of affairs to be attained by it, because that end can be produced by the action regardless of its motivation; so if the moral value of an action is to be connected to its motivation rather than its outcome, then it must lie "in the principle of the will without regard for the ends that can be brought about by such an action" (4:400), that is, in a moral principle that has nothing directly to do with the ends or consequences of the actions it commands. From this, Kant next infers, "duty is the necessity of an action from respect for law" rather than from any "inclination" – that is, naturally occurring desire – for an object or state of affairs (4:400). And from this – which is still supposed to be part of the common sense conception of a good will – the categorical imperative can be directly inferred: "Since I have deprived the will of every impulse that could arise for it from obeying some law," that is, every

LAWS OF FREEDOM

inclination for an object or state of affairs,

> nothing is left but the conformity of actions as such with universal law, which alone is to serve the will as its principle, that is, I ought never to act except in such a way that I could also will that my maxim should become a universal law.
>
> (4:402)

A "maxim" is the principle on which one actually acts, such as "I will enrich myself at all costs" or "I will never break a promise for reasons of self-interest," so this categorical imperative requires that each of us act only on principles on which *everyone* could act without contradiction: it requires that our "subjective principles of volition" also be "objectively valid" or universal laws (4:401n.). In my examples, the maxim "I will enrich myself at all costs" could *not* be acted upon by everyone, because something that I might do under that maxim is bound to conflict with something somebody else would do; but there would be no contradiction in all of us never breaking a promise for reasons of self-interest, so that could be a universal law and should be one. (I formulate this maxim as "I will never break a promise for reasons of self-interest" because there might be *other* reasons, such as saving an innocent life, that could make it permissible or even obligatory to break some promise.As can be seen from this, a maxim does not simply specify a general type of action to be performed or avoided, but also a specific reason for performing or avoiding that type of action.)

Kant's assumption that the fundamental principle of morality cannot be based on any mere desire for some end or object seems sound, but does his conclusion that this principle can therefore concern no end at all but only the universally valid form of our maxim in acting, that is, his purely formalistic conception of the categorical imperative, follow from this assumption? It does not seem to, since even if it is obvious that no object of merely contingent inclination could serve as the basis for morality, there still might be some sort of *necessary* object, perhaps of pure reason rather than inclination, which is the basis of the moral law, and if so then the fundamental principle of morality could be the substantive requirement to act only on maxims that would bring about that necessary object rather than just the formal requirement to act only on maxims that should also be universal laws. Let us look at Kant's further derivations of the categorical imperative to see whether he excludes this alternative or rather ends up exploiting it.

LAWS OF FREEDOM

Excerpted from *Kant*

In the second section of the *Groundwork*, Kant first argues against "popular moral philosophy" that the fundamental principle of morality can never be derived from examples of actual human conduct (as opposed to the imaginary examples or thought-experiments that he used in the first section, such as the case of the shopkeeper), because in real life people's innermost motivations are never certain, and are all too likely to turn out to be self-love, the "dear self" (G, 4:407). However, he claims that we can proceed by means of a philosophical analysis of the concept of a rational being instead of trying to extract our moral principle from examples of actual human behavior. In the first place, a rational being is one that acts, not just in accordance with laws (everything in nature acts according to some law, even stones falling in accordance with the law of gravity), but in accordance with its own *consciousness* or "representation" of laws (4:412). But to an imperfectly rational being, that is, one who has temptations to do otherwise than what its reason tells it to do, the laws in accordance with which it should act will present themselves as *constraints*, that is, "imperatives" (4:413).These imperatives can be of several different types.The major distinction between them is between those that are *hypothetical* and those that are *categorical,* that is, those that tell you what you must do if you want to attain some end – these are hypothetical – and those that tell you what you must do regardless of any such "reference to another end" – categorical imperatives (4:414). Hypothetical imperatives, in turn, can be divided into two further types: "problematic" ones, which tell you what you must do in order to attain some particular end you *might* have, and "assertoric" ones, which tell you what you need to do in order to attain an end you *do* have (4:415). Problematic hypothetical imperatives are obviously unfit to serve as principles of morality, since they clearly depend upon merely contingent ends. But assertoric hypothetical imperatives are also unfit to be moral principles, since the only end that everyone obviously does have is that of happiness, and that has already been excluded as a possible foundation for morality. Thus the only possible candidate for a fundamental principle of morality is a categorical imperative, one that tells you what you *must* do independent of any end you might have. Kant then argues that:

> When I think of a categorical imperative I know at once what it contains. For, since the imperative contains, beyond the law, only the necessity that the maxim be in conformity with this law, while the law contains no condition to which it would be limited, nothing is left with which the maxim of action is to conform but the universality of a law as such . . .

LAWS OF FREEDOM

> There is, therefore, only a single categorical imperative and it is this: act only in accordance with that maxim through which you can at the same time will that it become a universal law.
>
> (4:420–1)

Thus from the analysis of the concept of a rational being Kant ends up with the same imperative that he previously derived from the common-sense notions of good will and duty (with the possibly significant difference that the earlier formulation told us to act only on maxims that we *should* will to be universal laws while this one tells us to act only on maxims that we *could* will to be universal laws).

Is this argument any better than the earlier one? Actually, it looks worse, for not only does it again apparently simply overlook the possibility that in addition to the contingent ends that give rise to conditional, hypothetical imperatives, there might be a necessary end that could give rise to an unconditional, categorical imperative; it also simply assumes from the outset that a rational being must aim to act in accordance with a categorical imperative rather than merely hypothetical ones, and does not even attempt to derive this premise from anything like the commonly accepted conceptions of good will and duty appealed to in Section I.

The same apparently has to be said about Kant's derivation of the categorical imperative in the *Critique of Practical Reason*. Here Kant offers his most detailed account of why happiness cannot be the basis of a moral law: our conceptions of happiness are simply too indeterminate, for often what we think would make us happy at one moment conflicts with what we think would make us happy at another, or what one person thinks will make her happy conflicts with what would make another happy. (Kant relishes the irony in the story of Francis I of France and Charles V of the Holy Roman Empire, each of whom would have been made happy by the same thing, namely, possessing Milan. But obviously they could not both have Milan, so they could not both be happy in spite of agreeing on what would make them happy. See CPracR, 5:25–8.) So no genuine practical principle can be "material," or specify a particular object (5:21–2); instead, "If a rational being is to think of his maxims as practical universal laws, he can think of them only as principles that contain the determining ground of the will not by their matter but by their form," namely, that they have that form "by which they are fit for a giving of universal law" (5:27). But again, Kant simply assumes without argument that a rational being must will to act only in accordance with a truly universal law,

LAWS OF FREEDOM

and likewise that there are only contingent ends, no necessary end, so that the moral law must be strictly "formal" rather than "material."

When we return to the main line of Kant's argument in the second section of the *Groundwork*, however,we can see that the next thing that Kant does is precisely to fill the gap he has thus far left in his argument by overlooking the possibility of a necessary end by now introducing one. Kant does not, of course, acknowledge that there is a gap in his arguments to this point, but he seems to recognize that the purely negative arguments that he has offered thus far – arguments that arrive at the categorical imperative by the elimination of possible alternatives – would be more compelling if the principle were positively grounded in something of unconditional value. He acknowledges that "the principle of action being free from all influences of contingent grounds" needs to be connected "with the concept of the will of a rational being as such" (G, 4:426); in other words, precisely insofar as it is rational, a rational being needs a reason to adhere to a law, an end that can be advanced by and only by adherence to that law. And if the law is to be unconditionally valid, as the moral law is supposed to be, then that end must be unconditionally valuable. As Kant puts it, he must find "something the existence of which in itself has an absolute worth, something which as an end in itself could be . . . the ground of a possible categorical imperative." And then he goes on:

> Now I say that the human being and in general every rational being *exists* as an end in itself, *not merely as a means* to be used by this or that will at his discretion; instead he must in all his actions, whether directed to himself or also to other rational beings, always be regarded at the same time as an end.
>
> (4:428)

From this Kant derives the second main formulation of the categorical imperative: "So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means" (4:429).

Now if this imperative expresses the unconditional value of an end that can be the ground of any possible categorical imperative, then Kant's other formulations of that imperative, both the one commanding that we act only on universally valid maxims and any others to follow, ought to be derivable from it. So one question we need to ask is whether that is so. But before we can answer that question,we need to know just what this impressive-sounding statement means,

LAWS OF FREEDOM

and whether it can be proven any more convincingly than the original formulation of the categorical imperative. To determine what the statement means, we have to figure out what is meant by the concept of humanity as well as by the idea of an end in itself. One might think that by "humanity" Kant just means humankind, the biological species *homo sapiens*, or the defining characteristics of this species. In fact, Kant seems to mean something more like biological human beings insofar as they are also rational beings, and it is the embodiment of rational being rather than human life as such that he is declaring to be an end in itself. (In his Lectures on Ethics, Kant states that by engaging in various vices one can "throw away his humanity" without throwing away his life as such, and that "It is not life that is to be so highly treasured, but rather that one should live it throughout as a human being" [LEC, 27:341–2]. Kant did not believe in the sanctity of life as such.) Since human beings are the only rational beings we know, however, Kant often uses "rational being" and "humanity" interchangeably, and so we can glean what he means from statements about both. In the *Groundwork*, he says that "Rational nature is distinguished from the rest of nature by this, that it sets itself an end" (G, 4:437). A dozen years later, in the *Metaphysics of Morals*, he says that "what characterizes humanity (as distinguished from animality)" is the "capacity to set oneself an end – any end whatsoever" (MM, Doctrine of Virtue, Introduction, section VIII, 6:392), but also goes on say that "bound up with the end of humanity in our own person" there is that

> rational will, and so the duty, to make ourselves worthy of humanity by culture in general, by procuring or promoting the capacity to realize all sort of possible ends . . . In other words, the human being has a duty to cultivate the crude predispositions of his nature, by which the animal is first raised into the human being.

The term "humanity" in Kant's formula thus seems to mean our capacity freely to set ourselves ends – form intentions and adopt aims – and to entail a duty to develop the various abilities that as rational beings we can see will be necessary in order to pursue effectively and thus realize the ends that we have set for ourselves. It might seem as if effectively pursuing ends is different from setting them, thus that Kant's conception of humanity must consist of at least two parts. But in fact, a rational agent will not set ends for which he has no available means, so a failure to develop capacities to realize our possible ends in fact restricts our freedom to set ends. Kant's notion of humanity as the capacity to set ends freely thus contains an adequate basis for the duty to develop means to our ends as well.

LAWS OF FREEDOM

Excerpted from *Kant*

Now what can it mean to treat this capacity as an "end in itself," something that has "unconditional" or "absolute worth"? At the very least, something of unconditional value must not be destroyed or damaged for the sake of something of merely conditional value: thus our capacity to freely set ends is not to be sacrificed for the sake of any particular contingent end. Sometimes that seems to be all that Kant means, as when he says in the *Groundwork* that rational nature "must here be thought not as an end to be effected," that is, produced, but as an independently existing end, and hence thought only negatively, that is, as that which must never be acted against" (G, 4:437). But it is clear from Kant's remarks in the *Metaphysics of Morals* that there is more to making humanity our end than merely not acting against it; humanity includes capacities that must be developed in order to raise ourselves from the level of mere animality. Our humanity is both a predisposition and a potential, something that we must both preserve and promote. Further, although our humanity is something that is never to be sacrificed for any particular ends, it is nothing other than the capacity to freely set particular ends. Our humanity and our particular ends cannot simply be contrasted to each other, the latter simply being sacrificed for the former. Rather, the requirement that we make humanity our end and never merely a means requires that we set and pursue our particular ends in a way that is consistent with the preservation and promotion of our general capacity to set ends.

The capacity to set ends for ourselves sounds very much like the freedom that Kant talks about in his lectures on ethics: the capacity to set our own ends is freedom of choice, while freedom of choice in turn requires freedom of action because we cannot rationally choose ends that we have no way to pursue. In the lectures, as we saw earlier, Kant also says that freedom must be made "consistent with itself." What does that mean? One thing it seems to mean is that I must make free choices on particular occasions in a way that preserves and promotes my ability to make further free choices on other occasions. To use some of Kant's characteristic examples, particular decisions to commit suicide or get drunk considered by themselves would certainly be free choices – instances of setting myself "any end whatsoever" – but they would not be consistent with preserving and promoting my capacity to make further free choices: committing suicide, even if it is one free act, would obviously destroy me and therefore my ability to make any further free choices; choosing to get drunk, even if it is itself a free choice,would deprive me of the ability to make or successfully carry out free choices for some number of hours, and, were I to drive while drunk, could even end

LAWS OF FREEDOM

up killing me, thus directly destroying my freedom, or injuring or killing another thereby damaging or destroying her freedom.

Or I could kill someone else, and thus destroy his or her freedom – remember that Kant's requirement is that we treat humanity as an end and never merely as a means in my own person or in that of any other person. This means that my use of my own freedom on particular occasions must be consistent not only with my own future use of freedom but also with the preservation and promotion of the freedom of others. I could obviously make all sorts of choices that would be perfectly free choices, considered in isolation, and might even be consistent with my continued freedom, but which would be inconsistent with the preservation and promotion of the freedom of others. My decision to kidnap you might be a free choice, but would not be consistent with the preservation of your freedom; my decision not to pay my school taxes might be a free choice, but would not be consistent with the education of the children in my school district, thus with the promotion of their capacities to pursue their own freely chosen ends now or as they grow up. (Of course, we might expect or even hope that my violation of the freedom of others in such cases would lead to my punishment, and thereby a subsequent restriction or even destruction of my own freedom as well.) Consistently treating humanity as an end and never merely as a means requires the consistency of one's own free choices over time and consistency between one's own free choices and those of others both at one time and over time.The fundamental principle of morality commands that we seek such consistency in our use of freedom, and the concrete laws of morality are the more particular rules our reason tells us we must follow in order to achieve this general goal.

Interpreted along these lines, Kant's principle that we must always treat humanity as an end and never merely as a means not only sounds uplifting, but is also informative. But does it rest on anything more than mere assertion ("Now I say . . . ")? Does Kant have any argument for it?

At the outset of this chapter, I quoted Kant's early observation that even children are bitter at being constrained, and long to be able to make their own decisions. This might explain why one loves one's own freedom or humanity. But even if reflection on this fact about themselves were somehow to lead people to value everyone's freedom, the initial fact that even as children we love freedom seems to be only an anthropological or psychological fact, thus an empirical, contingent fact, not suitable for the foundation of a fundamental principle of morality, at least given Kant's expectation that such a principle must be valid for

LAWS OF FREEDOM

any possible rational being. In any case, in his mature publications on the foundations of morality Kant does not appeal to this psychological fact about us in order to justify the categorical imperative.

Kant says that his second formulation of the categorical imperative results from a step into metaphysics (G, 4:426), and some commentators have found in Kant a metaphysical argument, according to which the "conditional worth" or value that we assign to any particular end needs a foundation, indeed that it cannot simply be "relative" to some other conditional value but must ultimately be grounded in something of unconditional value, and that there is no other candidate for the unconditionally valuable source of conditionally valuable ends than our own capacity to choose those ends, so our capacity of choice must be the very thing that has unconditional value. But why shouldn't there be nothing but things of conditional or merely relative value, that is, things that are valuable only if something else is valued, but nothing that is of unconditional value? In fact, Kant does not suggest that the possibility of conditional value presupposes the existence of something with unconditional value; rather, he *assumes* that morality requires the existence of something of unconditional value, and *infers* from this that conditional or relative value cannot be the whole story about value. He does not try to infer the existence of unconditional value from the existence of conditional values (G, 4:428).

Are we in the end then just supposed to recognize the fundamental principle of morality as a basic norm that we all accept and which philosophy can clarify and confirm by deriving from it more concrete moral principles and duties that we all acknowledge, but which it cannot deduce from anything more basic? There is ample evidence to suggest exactly this. In the essay on metaphysical method written two decades before the *Groundwork*, Kant had said that the fundamental "material" principles of morality are "indemonstrable" (PNTM, 2:299). In the Preface to the *Groundwork* he had written that we "proceed analytically from common cognition to the determination of its supreme principle, and in turn synthetically from the examination of this principle and its sources back to the common cognition in which we find it used" (G, 4:392), which might be taken to mean that the only thing we can substantively add to the clarification of the supreme principle of morality is the confirmation of the correctness of our analysis of that principle by examples of its use. And in the *Critique of Practical Reason* he says that "consciousness of this fundamental law" is a "fact of reason" that just "forces itself upon us" (CPracR, 5:31). Maybe there can be no argument from some

LAWS OF FREEDOM

even more basic premise that there must be a fundamental principle of morality, although at least in the third section of the *Groundwork*, which we have not yet discussed, Kant tries to avoid this conclusion. But even if this is so, one could still argue that if there is a fundamental principle of morality, then it must have a certain character. Kant's second formulation of the categorical imperative might then be preferred to the first not because it has a better metaphysical foundation, but because it makes better sense of our common conception of our duties and it therefore better illuminates what the normative character of any moral law must be.

Perhaps in the end that is right. But there is one more thing that Kant says that we should think about. Back in his analysis of our common conception of the value of acting from duty as a motive, Kant had written that

> I cannot have respect for inclination as such, whether it is mine or that of another; I can at most in the first case approve it and in the second sometimes even love it, that is, regard it as favorable to my own advantage. Only what is connected with my will merely as ground and never as effect . . . can be an object of respect and so a command. Now an action from duty is to put aside entirely the influence of inclination.
>
> (G, 4:400)

This suggests that we can have no esteem or respect for what merely *happens* to us, but only for what we *do*, and if what we ultimately *do* is *choose* our ends and choose to develop and use various means to pursue them, but not in fact *realize* them, since that always depends at least in part on factors beyond our own action, then perhaps the only thing we can really respect is our choice of ends and the capacities on which that choice rests (just as the only thing we can really disrespect is a bad choice of ends, not the bad inclinations that people just happen to have or the bad things that just happen to them). This might suggest that humanity as the capacity to freely choose ends is the only candidate for something of *unconditional* value because as the essence of activity rather than passivity it is the only genuine object of *respect*.

Now it seems undeniable that the premise that we can have respect only for genuine actions is itself a normative assumption that is not derived from anything more fundamental, whether descriptive or normative. But perhaps some will find such a basic claim about moral *judgment* or *evaluation* more intuitively compelling than the more abstract and possibly unfamiliar theory of moral *value*

LAWS OF FREEDOM

that Kant enunciates in his principle that humanity should always be an end and never merely a means, and therefore find the former a possible premise for an argument to the latter. If not, well, then, Kant's argument is no worse off than before: it recognizes that concrete claims about moral norms can only be derived from something we acknowledge as a more fundamental moral norm, but that there can be no deduction of that fundamental norm from any metaphysical fact that is somehow more certain.We simply have to find what is presented as the most fundamental moral norm compelling, and certainly many people do find Kant's second formulation of the fundamental principle of morality immediately compelling.

Let us leave the problem of the derivability of Kant's second fundamental principle of morality there for now, and instead turn next to the question of whether Kant's other formulations of the categorical imperative can be derived from this one. After that, we can see whether even more concrete principles of duty can be derived from the categorical imperative, thereby lending it additional confirmation.

Universal law and humanity as an end in itself

Kant actually formulates the categorical imperative in at least five different ways, although he himself usually refers to only three (see G, 4:432, 436–7). Commentators have argued for every conceivable relationship among these formulations, but I will here develop the view that all the others may be derived from the formula of humanity as an end in itself (abbreviated "FHE"), in accordance with Kant's own suggestion that this formulation reveals the "ground of a possible categorical imperative."

What I have been referring to as Kant's second formulation of the categorical imperative is not in fact the first variant that follows his initial formulation, the formula of universal law ("FUL") requiring us to act only on maxims that we could also will to be universal laws (G, 4:421). Kant's first variant on that initial formulation is actually the formula of the universal *law of nature* ("FLN"), "act as if the maxim of your action were to become by your will a universal law of nature" (G, 4:421). Some commentators have claimed that this introduces something new into Kant's theory, namely a teleological conception according to which *nature* itself has certain purposes in giving us capacities and that we must act only in ways consistent with those purposes of nature. Kant's first illustration of this formulation is consistent with this interpretation: he argues that we should

LAWS OF FREEDOM

not commit suicide from *self-love* (that is, out of a desire to avoid further pain) because nature has given us the tendency to selflove to preserve our lives, not to end them (G, 4:422). But this teleological interpretation is not required by Kant's general conception of a law of nature, for all that Kant officially means by a law of nature is an unexceptionable uniformity in the behavior of some specified domain of objects: nature is just "the existence of things, insofar as that existence is determined according to universal laws" (PFM, §14, 4:294). So when Kant asks us by means of FUL to consider whether we could will a maxim on which we are considering acting to also be a universal law, or asks us to consider whether we could will to act upon our maxim if everyone else were also to do so, he is already asking us to consider whether we could will to act upon our maxim if that maxim were (somehow) to become one of the laws of nature in accordance with which everyone actually behaves, thus already implying FLN. Kant puts the same point in the *Critique of Practical Reason* when he says that

> The rule of judgment under laws of pure practical reason is this: ask yourself whether, if the action you propose were to take place by a law of the nature of which you were yourself a part, you could indeed regard it as possible through your will.
>
> (5:69)

Our actions take place in the natural world, so the question we are asking when we ask whether we could will our maxim as a universal law (FUL) is the same question as whether we could will it to be a law of nature (FLN).

Now as Kant points out, there are actually two questions I must ask when I ask whether I could will my proposed maxim to be a universal law of nature: first, whether it would even be logically possible for me to act on my maxim if everyone else were to do so too; and second, even if it would be logically possible for me to will the universalization of my maxim, whether that is something I could rationally will, that is, something that would be consistent with my willing things in a rational way (G, 4:424).What Kant means by the first of these tests is clear enough: if it would be impossible for me to act on my maxim if everyone did, then acting on my proposed maxim while willing it to be universal is logically impossible. For example, if everyone were to make false promises whenever they thought they could gain something by so doing, the very practice of promising – in which people act on promises made by others because they expect those promises will be kept – would quickly collapse, and once that happened it would be logically impossible

LAWS OF FREEDOM

for me to make even a false promise – the words "I promise" would be meaningless if there were no practice of promising based on the expectation that people generally keep their promises. The meaning of the second test is not quite so clear, but what Kant seems to have in mind is that the universalization of certain maxims would be inconsistent with a fundamental canon of rationality even if not logically impossible, namely the fundamental principle that if I am rationally to will an end then I must always be able to will an adequate means for it. As he puts it, "Whoever wills the end also wills (insofar as reason has decisive influence on his actions) the indispensably necessary means to it that are within his power" (G, 4:417). His idea would then be that while as a rational being you must will that there be suitable means available for your ends, whatever they might be, but that if you were to will the universalization of such maxims as "I will not cultivate my talents" or "I will not help others in need," that is, if you were to will that no one cultivates talents or helps anyone in need, then you would in fact be willing that adequate means for the realization of your ends not be available – the height of irrationality. Now Kant explicitly says only that the rule that if you will the end you must will some adequate means is the only principle of rationality needed to explain the force of *hypothetical* imperatives, e.g., such "rules of skill" as "If you want to assemble this furniture you must use a Phillips screwdriver," or such "rules of prudence" as "If you want to be healthy you must control your weight." This might make it seem as if this principle figures only in matters of prudence, not morality. But that does not follow, for Kant does not explain the moral, categorical imperative by this principle *alone*: the *moral* question is whether I could have adequate means for my ends *if I were to will the universalization of my proposed maxim* – as morality and only morality requires me to do. In other words, as the highest form of practical reason, morality comprises both the principle of universalization and the principle of instrumental rationality.

Kant associates his version of an important traditional distinction, that between perfect and imperfect duties, with the distinction between the two tests for universalizability. On Kant's account, perfect duties are those that prescribe a specific type of action, or more typically the omission of a specific type of action, while imperfect duties prescribe only a general goal or policy, but not the specific types of action by which that policy needs to be implemented. To use Kant's examples, suicide, or more precisely, in light of our previous discussion, suicide committed solely from the motivation of avoiding pain, is a specific type of action that is always prohibited, so the duty not to commit such suicide is a perfect duty;

LAWS OF FREEDOM

but, since you cannot possibly help everybody else in every way they might need, the general policy to help other people does not tell you what specific acts of beneficence to perform, and so is an imperfect duty. Kant's claim is that the proposed rejection of any perfect duty would fail the first test of universalizability, while the proposed rejection of any imperfect duty could pass the first test but would fail the second (G, 4:424). It is not clear whether this correlation holds in every case, but it is also not clear whether anything rides on that: as long as any duty that we are sure we have can be derived either from one or from both of the two parts of FUL / FLN, that would seem to confirm the adequacy of this version of the categorical imperative.

Of course, questions have been raised about whether FUL and / or FLN really do yield all our duties and only our duties. Many commentators have formulated immoral maxims that apparently pass the test of universalizability and clearly harmless ones that fail it, while several have argued that the universalizability test gives rise only to negative and not positive duties. The latter objection seems incorrect: if I must reject the maxims of letting all my talents rust or never helping anyone else, then I must accept their logical contraries, namely, maxims of cultivating at least some of my talents and helping at least some other people some of the time. To be sure, the latter maxims do not tell me specifically which talents I should develop or *which* people I should help *when, how*, and *how much* – but that is precisely the point that Kant himself makes by calling these maxims of imperfect duty, and if it is an objection at all then it would be an objection to the very idea of imperfect duty no matter how it was derived. But I do not want to go into these details here. For what I want to argue is that the force of the general idea of universalizability as a test of morality arises from the idea that humanity must always be treated as an end in itself (FHE), rather than FHE adding something to FUL / FLN, and if that is an adequate basis for all our actual duties then surely there must be a way to formulate FUL and / or FLN so that they are adequate as well.

The basic idea here is simply that FHE, the requirement that humanity whether in oneself or in anyone else must always be treated as an end and never merely as a means, requires that each one of us always respect the free choice and action of *everyone* else, and therefore act only on maxims that *could be accepted by everyone else* as preserving their capacity for free choice as well. In the first instance, that means that everyone else ought to be able to accept *my acting* on my proposed maxim, but full respect for their freedom also means that *they should be*

LAWS OF FREEDOM

able to adopt any maxim on which I propose to act, although they need not actually adopt every maxim on which I permissibly act. This is because to treat everyone equally as an end requires adopting only maxims on which everyone *could* act, if they were to so choose: there will be an unfair distribution of freedom, one on which not everyone is treated as an end in himself or some are treated as more of an end than others, if maxims are allowed on which some could act only if others cannot. Others will not be treating my humanity as equal in value to their own if they act on maxims that I could not also act on, and I will not be treating others as ends in themselves equal in value to myself if I act on maxims that they could not at the same time act upon. Kant puts the point in terms of ends – he says that to value others as ends and not merely as means requires that they "must also be able to contain in themselves the end of the very same action" I propose to do (G, 4:430) – but the same point goes for maxims: to treat others as ends equal in value to myself means that they must be free to adopt any maxim on which I propose to act. If they could not, then neither may I act upon such a maxim.

Of course, this means that treating everyone as equally free to exercise humanity or freedom of choice and action cannot be the same as anarchy: there will be many maxims we will all have to choose to forgo if we are all to treat each other as equally free. I obviously cannot adopt the maxim of committing homicide for any reason whatever if I value my own continued life and freedom, for that would mean allowing you to be free to act on the same maxim, and thus to kill me if you so choose. I cannot adopt the maxim of making false promises while allowing you the freedom of adopting the same maxim, for then I will not be able to accomplish anything at all by going through what would have become merely the motions of making a promise – again, in a world in which people routinely broke promises without good reason, no rational person would accept any promises, and thus the words "I promise . . ."would turn into meaningless noise. I cannot adopt the maxim of letting my talents rust if I am to allow you the same freedom, for then none of us might have the means necessary to realize any of our ends. And so on.Treating us all as equally free to adopt any maxim that any one of us is free to act upon means that we must all forgo certain maxims altogether and must all commit ourselves to adopting their contraries. That is why FHE implies FUL / FLN.

Confirmation of the categorical imperative from commonly recognized duties

Before we see how the imperative always to treat humanity as an end and never

LAWS OF FREEDOM

merely as a means also implies Kant's remaining formulations of the categorical imperative, let us stop to consider whether this formulation seems to be an adequate foundation for all the kinds of duties that we commonly recognize. This is not merely a natural question to ask, but also seems to be one that Kant himself promises to answer when he says, as we already noted, that we must be able to proceed "synthetically from the examination of this principle . . . to the common cognition in which we find it used" (G, 4:392). He illustrates both FUL / FLN and FHE with four examples, one each of a perfect duty to self, a perfect duty to others, imperfect duty to self, and imperfect duty to others, precisely because such a scheme is commonly recognized (G, 4:421–2n.). This classification is obviously exhaustive – leaving aside duties to God, which Kant rejects (see for example MM, Doctrine of Virtue, §18, 6:443–4) – so if Kant's formulations of the categorical imperative offer a way of grounding characteristic examples of duties in each of these four classes, that will be a strong argument from "common moral cognition" in their favor. As earlier noted, Kant's example of a perfect or strict duty to oneself is the prohibition of suicide. His argument is that one cannot "dispose of a human being in [one's] own person by maiming, damaging, or killing him" because one's humanity – not one's merely biological existence, but one's existence as a free and rational being capable of choosing and pursuing ends – is an end in itself; while to commit suicide, at least for such a reason as just to avoid further pain or disappointment, is to make "use of a person merely as a means to maintain a tolerable condition up to the end of life" (G, 4:429). The notion of making use of one's own existence merely as a means to achieving a certain condition in one's existence seems strange, but the general idea that one simply should not destroy something, namely, one's own humanity, that should always be treated as an end and never merely as a means, is clear enough. Presumably precisely the same argument applies in the case of homicide as well.

The permissibility of suicide was a standard topic in the ethical discussions of classical Stoicism and Epicureanism with which Kant was well acquainted, and had also become a fashionable topic in eighteenth-century Germany after the publication of Johann Goethe's bestseller *The Sorrows of Young Werther* (1774). For these reasons it greatly interested Kant – at least nothing that we know about him suggests that he ever struggled with any suicidal inclinations of his own – and he frequently discussed it. Two points that he raises elsewhere can help clarify his present argument. First, in the lectures on ethics that he gave in the years before publishing the *Groundwork*, he said that what is "inherently abominable" about

LAWS OF FREEDOM

suicide is "the fact that a man uses his freedom to destroy himself, when he ought to use it solely to live as a man"; a man is free "to dispose over everything pertaining to his person, but not over that person itself, nor can he use his freedom against himself" (LEC, 27:343). What this implies is, as I suggested earlier, that an act of suicide is itself a use of freedom, that is, a freely chosen act, but a free act against one's continued existence as a free agent, that is, one free act that would destroy the possibility of any further free acts. For that reason suicide cannot be endorsed but must be rejected in the name of humanity as freedom: what treating humanity as an end in itself requires is not that any free act considered in isolation, but that freedom as an ongoing condition, be preserved.

That we cannot allow any free act in isolation but must think instead of the preservation of freedom over a lifetime suggests that there is a certain quantitative aspect built into the requirement of treating humanity as an end and never merely a means, even though many people assume that quantitative considerations are relevant only to consequentialist theories such as utilitarianism. The second point that Kant makes about suicide in his lectures bears that out. Kant is generally inclined to treat the prohibition of suicide as absolute, but in pursuing the topic with his students he allows that certain exceptions may at least be possible. In particular, he discusses the case of the Roman leader Cato (Marcus Porcius Cato Uticensis, 95–46 BCE), who killed himself not to escape the tyranny of Julius Caesar personally but rather to encourage the Romans to "dedicate their final efforts to the defense of their freedom"(LEC, 27:370). Although Kant does not himself draw such a conclusion unequivocally, we can take this example to suggest that the (freely chosen) destruction of one free being in order to save many more free beings may be permissible, or even mandatory, because making humanity in both our own person and that of all others an end and never merely a means might well require preserving as many instances of humanity as possible; and in cases in which all instances cannot be preserved, then more rather than fewer instances should be preserved, even if it is our own instance of humanity that may have to be sacrificed in order to preserve others. Humanity is not just an abstraction, but something that exists in its instances, and so in making humanity our end numbers not only can but in fact must count. (However, Kant never suggests that making humanity our end requires *producing* more instances of humanity; he typically treats humanity, recall, as an end not to be acted against. Just why this should be so might not be easy to explain: it readily fits the ethical intuitions of those who believe the earth should not be overpopulated, for

LAWS OF FREEDOM

example, but not the religious views of those who believe they have a duty to procreate without limit.)

The same reasoning may apply in the case of homicide as well (which Kant does not actually discuss). Again, we may initially regard the prohibition of homicide as absolute, but in fact we do recognize exceptions to this prohibition.Thus,we acknowledge that the right to self-defense may sometimes license killing an attacker, and that means that we cannot think of the inviolability of each human life as if it were independent of all others, but rather recognize that sometimes one life can be preserved only at the cost of another, and that in certain circumstances one may have the right to preserve his or her own life rather than that of another. In this case, the reason for that right may be that one is innocent of any crime while one's attacker is not. But there will be other cases in which all the parties involved are equally innocent of any crime and yet they still cannot all be saved.To take one well-worn example, imagine that an out-of-control train is racing toward a switch where you just happen to be standing, and that a van with a family of six is stuck on the track to which the train will switch if you do nothing while a car with just one occupant is stuck on the other track.You might well think that it is not merely permissible but even obligatory for you to throw the switch so that only one person is killed by the train rather than six – your intervention will cause the death of the one, to be sure, but your decision to leave the switch as it is will cause the death of six, and that decision not to throw the switch would be just as much of an action on your part as your physical act of throwing the switch. If you accept this reasoning, you will be reasoning that if humanity is always an end, your duty is to preserve as many instances of humanity as possible, and that in unfortunate cases where for reasons beyond your own control not everyone can possibly be saved, then your duty is always to show your respect for humanity as an end in itself by saving more rather than fewer humans.

Thus, Kant's principle that humanity should always be an end and never merely a means can give a plausible derivation of our obligations in the prohibition of suicide as a perfect duty to self and the prohibition of homicide as a perfect duty to others. As I noted, Kant does not explicitly refer to the case of homicide; his example of a perfect duty to others is the prohibition of false promises, that is, promises made with no intention of being kept. (Not every *broken* promise is a *false* promise, since you may sometimes have morally permissible or even mandatory reasons for breaking a promise; a false promise is one that you *never meant* to keep.) In illustration of FUL / FLN, Kant had argued that making a

LAWS OF FREEDOM

false promise in order to accomplish some goal is impermissible because universalizing the practice of making false promises would undermine the practice of making promises altogether, and in that case you could not achieve your goal by making a false promise after all (G, 4:422). In illustrating FHE, Kant argues that in making a promise that you have no intention of keeping in order to accomplish a certain goal you are keeping your real intention and end hidden from the promisee, and thereby deceiving him into performing an action and adopting an end that he would not freely choose if he were properly informed about your real aim. False promises are impermissible, Kant concludes, because they "use the person of others merely as a means" to the hidden ends of the false promiser, "without taking into consideration that, as rational beings," the promisees "are always to be valued at the same time as ends, that is, only as beings who must also be able to contain in themselves the end of the very same action" (G, 4:430). That is, to treat others as ends and not merely as means is to treat them as entitled to *choose their own particular ends*, and thus to treat people as ends in themselves requires not merely preserving their *existence* as free beings but also preserving their capacity *to exercise their freedom* by choosing their own ends. Of course, this does not prohibit ever using another as a means at all, for even when you make an *honest* promise, say through a fair contract freely accepted by both parties, you are still using the other or the performance that the contract requires of him as a means for your own end in making the contract. But as long as the other party is agreeing to the contact *freely,* because he sees it as being in his own interest as well as in yours, then you are treating him as an end as well as a means, and this is what FHE requires.

So Kant's examples of perfect duties to self and others can plausibly be analyzed as duties to *preserve* the existence and the possibility of the *exercis*e of humanity, as the capacity to set ends freely. What about his examples of imperfect duties to self and others, which are prescriptions of certain general policies or goals rather than proscriptions of very specific types of actions. How can they be understood? Kant suggests that these should be understood as duties to *further* or *promote* humanity rather than to just *preserve* it. Now even though, as we have already seen, the duty to preserve humanity is in the first place a duty to preserve instances of humanity, by the duty to promote humanity Kant does not seem to mean a duty to *produce more instances* of humanity – he never asserted a duty to procreate. Rather, he has in mind duties to facilitate the realization of the particular ends that are freely chosen in the exercise of humanity

LAWS OF FREEDOM

both indirectly by the provision of general capabilities for successfully realizing such ends and directly by actually assisting in the realization of particular ends.The first of these cases is illustrated by Kant's example of an imperfect duty to oneself, namely, the duty to cultivate "predispositions to greater perfection," that is, skills and talents, in oneself (G, 4:430), because it is only by that means that one can develop the capacities that will be necessary to serve "all sorts of possible purposes" (4:423) that one may freely adopt over the course of one's life.We exercise our humanity precisely by freely choosing ends, but the failure to develop adequate means to our ends restricts our choice of ends. It is therefore part of treating humanity as an end to take steps to promote the effectiveness of those choices.This is not a prudential or utilitarian argument that we will be *happier* if we take steps to enable ourselves to realize more rather than fewer of our chosen ends – though no doubt we usually will be – but is rather an argument that because our free choice of ends is an intrinsically valuable exercise of our humanity and cultivating our talents in order to realize these ends is necessary to maximize our freedom to choose ends, cultivating those talents is also part of what is required to treat humanity in our own person as an end in itself.

Before turning to Kant's example of imperfect duty to others, one observation about this imperfect duty to oneself is in order. At one point, Kant says that "as a rational being [one] necessarily wills that all the capacities in him be developed" (G, 4:423). This cannot be true, because in many cases it simply will not be possible to develop all of one's potential skills or talents. One might have equal potential to become a great violinist or a great linebacker, but it is extremely unlikely that one could actually become both, because of the amount of practice time each would require, the incompatible developments in physique they would require, and so on. Usually one will have to make a choice of which talents to cultivate, and factors other than the completely general obligation to cultivate some talents will be necessary to make that choice. Again one such factor might be quantitative – one might ask which skill will ultimately allow one to realize more of one's possible ends, or even more of one's own ends as well as the ends of others whom one might help through one's own talents and their fruits. Happiness too might be a factor – faced with two equally good ways of facilitating your successful pursuit of "all sorts of possible purposes," you might simply ask yourself which one would make you happier. Of course, as Kant likes to stress, we are not particularly good at answering that question for ourselves. But our ultimate question must always be, are we maximizing rather than compromising our

LAWS OF FREEDOM

freedom to set ends?

Finally, Kant's explanation of the duty to assist others in the realization of their ends also turns on the assumption that to treat humanity as an end and never merely as a means requires treating the ends that other people choose in the exercise of their humanity as worthy of promotion because of the value of their humanity. Merely preserving the existence of others or even not directly interfering with their freedom of choice is not enough; as Kant says:

> there is still only a negative and not a positive agreement with humanity as an end in itself unless everyone also tries, as far as he can, to further the ends of others. For the ends of a subject who is an end in itself must as far as possible be also my ends, if that representation is to have its full effect in me.
>
> (G, 4:430)

Again, the argument is not a utilitarian argument: the claim is not hat I should assist others in the realization of their goals because that will make them happier, though no doubt it usually will. The duty must rest on the value of freedom, not of happiness. Although Kant does not say so explicitly, it looks again as if his best argument would be that making means for the realization of their ends available to others when they cannot do so themselves actually increases the scope of their freedom of choice and respects their humanity in this way.

Now, of course, we will want to recognize at most a duty to promote the *morally permissible* ends of others. But this is readily explained on Kant's analysis: morally impermissible ends would be those that would in some way destroy or violate humanity, whether in the person whose ends they are or in others, and we obviously have no duty to assist in that. On the contrary, since our duty to assist in the realization of the particular ends of others derives from our general duty to preserve and promote humanity, we can have such a duty only when those particular ends are themselves consistent with that general duty. Kant also observes later that "it is open to me to refuse" to help others with "many things that they think will make them happy but that I do not, as long as they have no right to demand them from me as what is theirs" (MM, Doctrine of Virtue, Introduction, Section V, 6:387) (that is, as long as I do not already we them what they want because of some prior contract, promise, etc.). This reservation could easily be explained if our duty were simply to promote the happiness of others – of course we all have to exercise our own judgment in figuring out how to fulfill

LAWS OF FREEDOM

our duties. The explanation will have to be more subtle given that Kant's underlying theory is not that happiness is intrinsically valuable and that we have a duty toward the happiness of others for that reason. His thought must rather be that even though we have a general duty to assist in the realization of the ends of others, it is of course impossible for us to assist with the realization of all the ends of all other people; so we must again appeal to further factors in deciding where to address our necessarily limited assistance. At this point it certainly seems appropriate to appeal to such considerations of number, reliability, and efficiency: how can we help the *most* other people? How can we *most reliably* help others? How can we *most effectively* help other people? That is, how can we most effectively and extensively expand the freedom of others to set their own ends? In trying to answer these questions, we will certainly have to make our own judgments about what is actually in the best interest of those whom we would try to help.

Kant's derivation of specific examples of duties from the general requirement that we treat humanity as an end and never merely as a means thus seems plausible. I will just add one remark before returning to the remaining formulations of the categorical imperative. Kant offers the duty of perfecting one's own natural predispositions and assisting in the realization of the ends of others merely as examples of imperfect duties to oneself and to others respectively. But in the later *Metaphysics of Morals,* he will argue that one's only duty to oneself is to promote one's perfection and that one's only duty to others is to promote their *happiness*, thus that one has no duty to promote one's own happiness or the perfection of others (MM, Doctrine of Virtue, Introduction, Sections IV–V, 6:385–8). His stated reasons for these claims are, first, that one can have a duty only to do something to which one is not naturally inclined, but everyone is naturally inclined to pursue their own happiness, so one cannot have a duty toward that, and, second, that the perfection of humans consists precisely in their setting their ends in accordance with their own concepts of duty, and obviously no one can do that for someone else (6:386). Both these arguments are weak. First, while one may not need to constrain oneself to pursue some immediate inclination, one's long-term happiness often conflicts with immediate inclination, and one may well need to constrain oneself to pursue it. So one's long-term happiness may often seem more like a duty than an inclination; and if we have a duty to promote the long-term happiness of *others* because of the value of their humanity, then we could well have a duty to promote our own *long-term* happiness because of the value of our

LAWS OF FREEDOM

own humanity. Second, while we certainly cannot make each other's choices, Kant's discussion of self-perfection ultimately makes it clear that this involves far more than simply making choices in accordance with duty: it involves the perfection of a whole variety of natural as well as moral capacities that we need in order to make wise choices, whether moral or just prudential, as well as to realize them successfully. In other words, self-perfection requires the education of our natural and moral capacities, and we can certainly assist others with that. For example,we can have a duty to assist in the education of children, both our own and those of others. Such a duty may be in part a perfect duty – our obligations to pay our school taxes and make sure our own children go to school until they are may be specific and unremitting – but it may at least in part be imperfect – there may be all sorts of ways in which we should promote the education of children, whether our own or others', that cannot be specified in such precise ways. But again, our underlying obligation toward others will be to help them increase the scope for the exercise of their freedom, not merely to promote their happiness, even though the latter will often be the result of the former.

Autonomy and the realm of ends

Let us now return to Kant's further formulations of the categorical imperative. He twice speaks of a third formulation, after FUL / FLN and FHE, but each time he mentions a different formulation. So there seem to be two further formulations, not identical but presumably related. Kant's first derives the "third practical principle of the will" from the preceding formulations of the categorical imperative thus:

> The ground of all practical lawgiving lies (in accordance with the first principle) objectively in the rule and the form of universality which makes it fit to be a law (indeed a law of nature); subjectively, however, it lies in the end; but the subject of all ends is every rational being as an end in itself (in accordance with the second principle); from this there follows now the third practical principle of the will, as supreme condition of its harmony with universal practical principle, the idea of the will of every rational being as a will giving universal law.
>
> (G, 4:431)

A page later he gives a slightly different formulation when he says that "the principle of every human will as a will giving universal law through all its maxims"

LAWS OF FREEDOM

(G, 4:432).Together, these two statements suggest that the third formulation of the categorical imperative is something like "Act only on maxims that could be given by all human wills as part of a *complete system* of maxims." Kant calls this third formulation "the principle of the autonomy of the will in contrast with every other, which I accordingly count as heteronomy" (G, 4:433), so this version is often called the formula of autonomy (FA). His reason for this name is his definition of "autonomy" as "the property of the will by which it is a law to itself (independently of any property of the objects of volition)" (G, 4:440). His idea is that for your will to be determined simply by inclination toward some object is for your will as it were to allow itself to be pushed around by those inclinations, or to be "heteronomous," rather than to be freely self-determined, or "autonomous," and that the only way for your will to be free or autonomous is for it to be governed by a law that it gives itself rather than to allow itself to act on whatever mere inclination happens to be alluring at the moment. And because your will would be determined heteronomously rather than autonomously whether it let itself be pushed around by one of your own inclinations or by someone else's inclination (perhaps the latter would be the everyday sense of heteronomy), the only rule that can truly free you (along with everyone else) from heteronomy and truly realize your potential for autonomy is the rule that *no one* should act on any maxim determined by mere inclination, but rather that *all* should act only on a set of rational principles consistent with the freedom of each, thus a system of maxims that each could freely will. It may seem strange that the freedom of anyone can be realized – preserved and promoted – only if all act on a common system of universalizable maxims, but Kant's idea is that if that is not the case, then someone will always be pushed around by some mere inclination, whether his own or someone else's.

Kant's claim, then, is that the formula of autonomy (FA) follows from FUL / FLN and FHE because treating every human being as an *end in itself* requires that *all* of the maxims on which you act could be freely willed by *all* human beings, and that only if all act on such a set of maxims will the freedom of all be preserved and promoted in the way commensurate with the value of each person as an end in itself.As we earlier observed, however, FUL / FLN itself follows from FHE: the requirement to treat humanity whether in yourself or in anyone else as an end in itself already requires that each of us act only on maxims that could be freely accepted by everyone else; so we can also see FA as following from FHE alone.

Kant next says that "the concept of every rational being as one who must regard himself as giving universal law through all the maxims of his will . . . leads

LAWS OF FREEDOM

to a very fruitful concept dependent upon it, namely that of a realm of ends," where by such a realm he understands "a systematic union of various rational beings through common laws," or more fully "a whole of all ends in systematic connection (a whole both of rational beings as ends in themselves and of the ends of his own that each may set himself)" (G, 4:433). He then represents the principle "that all maxims from one's own lawgiving are to harmonize with a possible realm of ends, as with a kingdom of nature" (the formula of the realm of ends, or FRE) as the third formulation of the categorical imperative, instead of FA, when he derives it, just like FA, as the "complete determination of all maxims" following from the prior requirements that all maxims have "a form, which consists in universality," stated in the formula that "maxims must be chosen as if they were to hold as universal laws of nature," and "a matter, namely an end," stated in the formula "that a rational being, as an end by its nature and hence as an end in itself, must in every maxim serve as the limiting condition of all merely relative and arbitrary ends" (G, 4:436). The idea behind the derivation of FRE from FUL / FLN and FHE should be immediately clear from our original analysis of what Kant means by treating humanity as an end in itself: it is just that since to treat *any* human being as an end in itself is both to preserve that person's existence as a being capable of freely setting ends, which they can do only if they can effectively pursue the ends they would choose, to treat all human beings as ends in themselves is both to preserve the existence and freedom of all such beings (or as many as possible) "as a whole" in "systematic connection" *and* to promote the realization of as many as possible of their freely chosen ends as a "whole" in "systematic connection" – thus, to act only on maxims consistent with a realm of ends and indeed to work toward the realization of such a realm. Once again, of course, since FHE itself already implies FUL / FLN, FRE can be seen as really following from FHE alone.

Kant's moral theory is often described as "non-consequentialist," as if it took no account of the consequences of our actions, but that is clearly misleading. To be sure, his theory gives no intrinsic value to states of affairs or consequences merely because they are desired as objects of inclination, but it greatly values the ability to realize our freely chosen ends as a necessary condition of our capacity for free choice itself. The realm of ends as the systematic union both of human beings as ends in themselves and of their freely chosen particular ends would be nothing other than the consequence of everyone's acting on the categorical imperative; and while the idea of humanity as an end in itself may best express the ultimate *source* of value in Kant's moral theory, the idea of all humanity as a kingdom of

LAWS OF FREEDOM

ends may best express the ultimate *consequences* of acknowledging this value, and thus give us our clearest idea of the *goal* or what Kant calls "object" of morality. The full force of Kant's idea of the realm of ends is often understated when it is described, for example by John Rawls, simply as the idea of a "moral commonwealth" in which we are all co-equal "legislators . . . of the public moral law." This brings out the first half of Kant's idea – that morality requires us to think of every person as equally free and thus as an equal legislator of the maxims on which we must all act – but does not bring out the second half of Kant's idea – that morality requires us to promote the systematic realization of freely chosen particular ends because we cannot rationally choose ends we cannot realize. Allen Wood recognizes that "Rational beings constitute a *realm* to the extent that their ends form a *system*" in which "these ends are not only mutually consistent, but also harmonious and reciprocally supportive," thus that "the laws of a realm are such that universally following them would result in the agreement and mutual furthering of the ends of all rational beings in a single unified teleological system." In spite of this, he also holds that "FA and FRE are merely general characterizations of the entire system of moral laws, which resist direct application to individual cases," and that we can only decide individual cases by applying all of FUL / FLN, FHE, FA and FRE to particular cases.The view I have presented here is that FHE tells us in the most basic terms how we must treat people in order to be moral; that FUL / FLN and FA successively bring out the universalistic implications of FHE, FUL / FLN telling us first that we must treat each of our maxims as universally acceptable and FA then telling us that we must treat the system of all of them as such; but that only FRE fully brings out FHE's implication that we must act so that not just human beings but also their freely chosen ends can become a systematic union. It, therefore, provides Kant's most concrete and fullest account of the goals of moral conduct.

Having completed his formulations of the categorical imperative, Kant tells us that he has only analyzed or explicated "the generally received concept of autonomy" and not yet "affirmed its truth" (G, 4:444). In other words, Kant has not in fact given up on the idea of proving that we are subject to the moral law by more than just an appeal to common sense. To do that, however, or to show that "morality is no phantom," he says, "requires a possible synthetic use if pure practical reason" (G, 4:445), which he will provide in the final section of the *Groundwork*. This section introduces Kant's theory of the freedom of the will into his moral philosophy, because he holds that we can only realize our freedom by acting in

LAWS OF FREEDOM

accordance with the moral law but can only act in accordance with the moral law if we are free, thus we must prove that we have freedom of the will if we are to prove both that we ought to obey the moral law and that we can. But Kant's views on the freedom of the will are complex, even paradoxical, and underwent considerable evolution over his career.They deserve a chapter of their own. Before we see how Kant more fully developed his idea of a realm of ends into the form of the system of political and ethical duties that he finally published, a dozen years after the *Groundwork* and at the very end of his career, in the *Metaphysics of Morals,* we must therefore pause to discuss Kant's views on the freedom of the will and two other "postulates of pure practical reason" that he often links to that topic, namely the postulates of immortality and the existence of God.

Summary

Kant begins his presentation of his normative ethics in both the *Groundwork* and the *Critique of Practical Reason*, and thus his account of autonomy in its practical sense, with the formulation of the categorical imperative that we must act only on maxims that we could also will to be acted upon by everyone else. In the *Groundwork*, he then goes on to formulate the principles that we should act only on maxims that treat humanity in both ourselves and others as an end in itself of unconditional value, never merely as a means, that we should act only on maxims that could be universally legislated within a consistent system of maxims, and that we should act so as to bring about a realm of ends, in which each human being is treated as an end in him- or herself and his or her freely chosen ends are promoted to the extent that so doing is consistent with treating each as an end in him- or herself. I have argued here that Kant's most fundamental normative notion is the idea of treating humanity as an end in itself, that is, treating each human being as an autonomous agent capable of setting his or her ends both freely and yet in harmony with others, and that the other formulations of the categorical imperative as well as Kant's examples of the chief classes of moral duties can all be derived from this basic idea. Now we are to see how Kant attempts to prove that this conception of the requirements of morality is binding on us.