

3

DOING HARM

3.1 Deontology

Consequentialists accept a simple theory of the right. There is one and only one factor that has any intrinsic moral significance in determining the status of an act: the goodness of that act's consequences (as compared to the consequences of the alternative acts available to the agent). Of course, as we have also seen (in 2.6), consequentialists can find a place for the more familiar rules of commonsense morality as well—a requirement to keep one's promises, a prohibition against lying, and so on. These secondary rules pick out types of acts that generally have good or bad consequences, and so provide helpful guidance in normal deliberation. But whatever the usefulness of these more familiar rules, the fact remains that from the consequentialist point of view they have no intrinsic moral significance. The various factors identified by these rules (promise keeping, truth telling, and so on) may be instrumentally valuable in discovering which act *has* the best results; but in and of themselves they play no role in making it be the case that a given act has the moral status that it does. Ultimately, an act is right if and only if it will have the best results; morally speaking, nothing else matters.

So say the consequentialists. But many people find this a difficult position to accept. Intuitively, at least, it seems that many normative factors have a moral significance that is not exhausted by the good or bad results that they normally involve. Intuitively, that is, most of us believe that there are other factors that have intrinsic moral significance beyond that of goodness of outcomes. Of course, as I have repeatedly emphasized, virtually no one denies that goodness of outcomes is *one* of the intrinsically relevant factors; but the point to remember now is that one can accept this modest claim while still rejecting the consequentialist's considerably bolder claim that this is the *only* factor with intrinsic significance.

In opposition to consequentialism, most people believe that there are *several* normative factors with intrinsic moral significance. And the moral status of an act is determined by the interplay of all of these various factors—not good consequences alone. In principle, at least, this means that a given act might be morally forbidden even though it has the best results. For there may be a second normative factor that comes into play—a factor capable of “outweighing” or “overriding” the fact that the act will have the best possible results overall.

An example should make the point clear. Imagine that there are five patients, each of whom will soon die unless they receive an appropriate transplanted organ: one needs a heart, two need kidneys, one needs a liver, and the fifth needs new lungs. Unfortunately, due to tissue incompatibilities, none of the five can act as donor for the others. But here is Chuck, who is in the hospital for some fairly routine tests. The hospital computer reveals that his tissue is completely compatible with the five patients. You are a surgeon, and it now occurs to you that you could chop up Chuck and use his organs to save the five others. What should you do?

If you are a consequentialist, the answer depends only on the goodness of the results. Of course, until we specify a particular theory of the good, we can't say for sure which results would be better. But for simplicity, let's consider the answer of the utilitarian—the consequentialist who accepts a welfare account of the good. (Other plausible theories of the good are likely to give the same answers in this case, so restricting our attention to utilitarianism shouldn't beg any questions against consequentialism. But see 6.4.)

Your choices are these: do nothing, in which case five people will die and one person will live; or chop up Chuck, in which case Chuck will obviously die but *five* people will live. From the utilitarian standpoint the results certainly seem to be better if you chop up Chuck. After all, if everyone counts equally, then it is simply a matter of five versus one. Obviously, it is a horrible result that Chuck will end up dead; but it would be an even worse result if *five* people end up dead. So the right thing to do—according to utilitarianism—is to kill Chuck.

But this seems absurd. Intuitively, at least, most of us have little doubt that it is morally forbidden to chop up an innocent person, even if this is the only way to save five other innocent people from death. Some acts are morally off-limits; they are forbidden, even if the results would be good. So there must be other factors that can play a role in determining the moral status of an act—beyond the goodness of that act's outcome. In the case at hand, it seems highly relevant that the admittedly good result of saving five people can only be brought about by an act that involves harming, indeed *killing*, an innocent person!

Here, it seems, we have identified a second relevant normative factor. The moral status of an act depends not only on the goodness of its results,

but also on whether or not it involves doing harm to someone. (By itself, perhaps, this case cannot establish whether we care about harming per se, or only killing in particular; but similar examples would quickly support the former conclusion.) What's more, the organ transplant case suggests that most of us believe that this second factor straightforwardly outweighs the first factor, good results: even if harming someone is the only way to bring about the best results, it just can't be done. Morally, it is more important to avoid doing harm than it is to *save* people who need our help. This, at any rate, appears to be the view of commonsense morality.

It is important to be clear about what this view holds. It does *not* claim that it is morally irrelevant that five people are going to die. The possibility of saving five lives remains a significant factor in determining what we should do. Obviously, if there were some way to save the five without killing Chuck then we would be morally required to do so. But commonsense morality insists that the possibility of saving five lives—even though this is the best result overall—is not the only factor of intrinsic moral significance. It is also relevant that the only way to bring about the best results involves *doing* harm. And given a conflict between these two factors—given the necessity of either doing harm or failing to produce the best possible results—it is more important to avoid doing harm.

Now in principle, a moral theory could be *pluralistic*—recognizing more than one normative factor—without going so far as to hold that goodness of results can sometimes be *outweighed* by these other factors. (For example, a moral theory might claim that doing harm is intrinsically relevant but only as a tie breaker: according to such a theory, if two acts would have equally good results, it would be morally preferable to pick the act that does not involve doing harm; but if doing harm was necessary to bring about *better* results, it would still be permissible.) So it is important to stress that commonsense morality appears to do more than merely recognize the intrinsic significance of harm doing: it invests this factor with considerable weight—a force sufficient to outweigh good results. We are forbidden to do harm, even if this is the only way to bring about the best results overall. Indeed, the significance of harm doing is so great—according to commonsense moral intuition—that all other things being equal, if an act involves doing harm, it is simply forbidden.

We can express this point by saying that commonsense morality recognizes a *constraint* against doing harm. That is to say, there is a prohibition which rules out performing a certain type of act—in this case, doing harm—even when performing an act of that type is necessary to bring about the best results overall. Now obviously enough, consequentialist theories leave no room for moral constraints, since such theories hold that whatever act will lead to the best results is always permissible. But it is precisely at this point that commonsense morality diverges from consequentialism. Let us call any moral theory that does incorporate constraints *de-*

ontological. It appears, then, that commonsense morality is deontological, rather than consequentialist.

Unfortunately, although the term “deontology” is widely used in contemporary moral philosophy, there is nothing like a standard or received definition of the term. In defining deontology, I have appealed to the concept of a constraint: deontologists, unlike consequentialists, believe in the existence of constraints, which erect moral barriers to the promotion of the good. But why not take the more straightforward approach and simply say that deontologists are all those who reject consequentialism? This would make for a sharp contrast between deontology and consequentialism, but it would also be far too broad to capture anything like the contemporary philosophical use of the term “deontology.” (A moral system in which every act is permissible would certainly not be consequentialist—but it would never be classified as deontological either.)

Similarly, if we hope to capture contemporary usage, it won’t quite do to label as deontologists all those who accept additional normative factors, beyond that of goodness of results: we must add the further stipulation that in at least some cases the effect of these additional factors is to make certain acts morally forbidden, even though these acts may lead to the best possible results overall. In short, we must say that deontologists are those who believe in additional normative factors that generate constraints.

This characterization of deontology—in terms of constraints—is by no means the most common one. I believe, however, that it comports reasonably well with contemporary use. (And, at the very least, it is probably sufficient for the purposes of recognizing typical deontological positions.) In any event, other common attempts to define deontology suffer from their own difficulties as well.

For example, it is sometimes suggested that in contrast to consequentialism—which assesses the moral status of an act solely in terms of its effects—deontological theories also assess an act in terms of its *intrinsic character*. But it is difficult to see how to come up with an account of “intrinsic character” so that this is necessarily true. (If I harm someone, generally it is not the *intrinsic* character of the act that makes it harm doing, but rather the fact that the act has a certain kind of result.) Similarly, it is sometimes suggested that deontologists, unlike consequentialists, believe in absolute rules. But this too is an inadequate way of drawing the contrast. As we will see, deontologists need not actually be absolutists about their rules; and, at any rate, consequentialists *are* absolutists: they are absolutists with regard to the requirement to do the act with the best results.

Another proposal is that deontologists—in contrast to consequentialists—believe in “the priority of the right over the good.” The idea behind this expression is this. Consequentialists start with a theory of the good, and then “define” the right in terms of it: the right act is simply that act that leads to the best results. For consequentialists, then, the good is prior to the right. In

contrast, deontologists believe that sometimes the right act is *not* the one that leads to the best results overall. Accordingly, the right cannot be simply defined in terms of the good or “reduced” to the good; it has some content independent of the good, and is in this sense “prior” to it. Understood in this way, the claim that deontologists believe in the priority of the right over the good is merely a shorthand way of expressing this thought: deontologists believe that goodness of results is not the only normative factor relevant to determining the moral status of an act; other factors are relevant as well, and these can have the effect of making it impermissible to perform the act with the best results. So understood, this is more or less equivalent to the way in which I have characterized deontology.

But the claim that deontologists—unlike consequentialists—accept “the priority of the right over the good” is potentially also a rather misleading one. For it may create the impression that somehow it is only deontologists who recognize the importance of doing the right thing—whereas the reality, of course, is that consequentialists believe in this as well. Talk of the priority of the right over the good may also create the impression that deontologists must *define* the good in terms of the right (just as consequentialists are said to “define” the right in terms of the good). But this certainly need not be the case: the deontologist need not believe that the right is *prior* to the good in any robust sense; it is simply that the deontologist—unlike the consequentialist—holds that considerations of goodness do not exhaust the considerations relevant to determining the act’s rightness. In particular, certain factors may make it morally impermissible to perform an act, even though that act would have the best results. It is only in this sense that deontologists must believe that the right is “prior” to the good.

For all of these various reasons, then, I think it preferable to define deontology in the way that I did—that is, in terms of the acceptance of constraints. But however we define deontology, the important point is really this: commonsense morality does indeed accept constraints; and this is the reason why most of us find consequentialism an inadequate moral theory.

Here is another way to express the point. Admittedly, consequentialists can find *some* place in their theory for the rules of commonsense morality, including a rule against harming people. (After all, in *normal* situations harming an innocent person is hardly likely to bring about the best results.) But it doesn’t follow that consequentialists can give these rules the same sort of weight that they have in ordinary moral thinking. And what the organ transplant case suggests is that consequentialists fail to give these rules sufficient weight; they permit breaking the rules in cases where the rules should not be broken.

For the consequentialist, the familiar rules of ordinary morality are merely “rules of thumb”: they pick out acts that in normal circumstances tend to have good or bad results; the rules should therefore be followed *all other things being equal*. But in exceptional cases, in cases where obeying

the rules would not actually lead to the best results overall, the rules have no further weight or significance—they are, after all, mere rules of thumb—and so the consequentialist thinks they should be broken. In contrast, what most of us believe is that these rules identify further factors that have a direct role to play in determining whether an act is right or wrong. And if these factors do have intrinsic weight—if they matter in their own right and not merely as indicators of good or bad results—then the mere fact that breaking some rule will lead to the best results overall will not necessarily be sufficient, in and of itself, to justify breaking the rule.

Other examples reinforce the point. Imagine that some horrible crime has been committed, and unless the guilty party is found soon, an angry mob will riot in the streets, killing several innocent people. You are the sheriff, and you cannot figure out who actually committed the crime. But it occurs to you that you are in a position to secretly fabricate evidence, framing an innocent person, thereby satisfying the mob and saving the lives of those who would otherwise be killed during the riots. What should you do? It certainly looks as though the consequentialist must classify this as another “exception”—a case in which the familiar rules of justice are to be violated. But intuitively, at least, most of us would find it morally unacceptable to knowingly frame an innocent man, even if this is the only way to save lives. So here too we find consequentialism endorsing the breaking of rules in a case where they should not be broken. (Presumably this is the thought that lay behind the attempt—noted above—to define deontology in terms of absolutism; I think it is a mistake, however, to focus on absolutism *per se*, rather than on the particular rules that the consequentialist is prepared to break.)

What the organ transplant case, the angry mob case, and similar examples seem to show is that consequentialism is inadequate as a moral theory. Because it accepts an inappropriately short list of normative factors, it permits acts that are not in fact morally permissible. It tells us to chop up Chuck, to frame the innocent man, and so on, even though these are clearly immoral acts. In short, consequentialism *permits too much*. At any rate, that’s the objection offered by consequentialism’s deontological critics.

There are, however, a number of ways in which consequentialists can try to resist this objection. Obviously enough, the objection gets its bite from the claim that in various cases consequentialism yields intuitively unacceptable results. Accordingly, one way to undermine the force of the objection is to deny that this is the case, that is, to deny that consequentialism actually supports performing the intuitively unacceptable acts.

In the organ transplant case, for example, the problem is that consequentialism appears to support chopping up Chuck, an act that certainly strikes us, intuitively, as morally impermissible. But the consequentialist could insist that—initial appearances to the contrary notwithstanding—killing Chuck isn’t likely to have the best results. After all, organ transplants have

a very high failure rate; if you chop up Chuck you could easily end up with *six* people dead rather than merely five. In light of the risks, it is far from obvious that killing Chuck will do more good than harm. Furthermore, it is important to remember that *all* the results have to be taken into account—not only the short term benefits but the long term costs as well. Consider, then, the likely effects of your act once it comes to light that a perfectly healthy person was killed during a routine visit to the hospital: people will start thinking twice about seeing doctors or going to hospitals, and in the long run the negative effects on public health will be horrible. The total outcome will thus be much worse if you kill Chuck than if you refuse to do it and simply let the five original patients die. But if chopping up Chuck isn't really the act with the best results, then consequentialism doesn't endorse performing that act, and so consequentialism hasn't been shown to be intuitively unacceptable.

A reply along these lines obviously relies on its various empirical claims, and so the critic of consequentialism might try to challenge them. But there is another possibility as well. The critic could simply alter the example, attempting to arrive at a version of the case in which it is undeniable that chopping up Chuck *would* have the best results overall. For example, we might be asked to imagine a time in which organ transplant procedures have been perfected, thus guaranteeing that the five will indeed be saved if only Chuck's organs become available. Similarly, to avoid the potential long term negative effects of your being open about having killed Chuck, we might be asked to imagine that you find a way to kill him secretly, making it look as though he died of a rare and unexpected medical complication. Properly modified, it seems as though we will eventually arrive at a version of the organ case in which it is indeed true that killing Chuck so as to save the five patients will have the best results overall.

Some consequentialists will still deny this. They might claim that for no version of the story will it be true that killing Chuck has the best results overall. Of course, the more we add to the story—the more unrealistic we make it—the harder it is for the consequentialist to insist that the results of killing Chuck simply *cannot* be best. But this then points to a slightly different reply open to the consequentialist: he might agree that *eventually* we may be able to tell a version of the story in which killing Chuck has the best results; but he will insist that once we actually produce such an example—filling in the details so that it is clear that anything other than killing Chuck will have even worse results overall—we will no longer find it counterintuitive to claim that (given all the bizarre circumstances of the case) killing Chuck is the right thing to do.

Other consequentialists may prefer a somewhat different tack. They might insist that for all *realistic* examples, consequentialism yields intuitively acceptable results. They may concede that for sufficiently outlandish, “science fiction”-type cases, consequentialism may well support acts that

seem intuitively unacceptable. But—they will insist—for such highly unrealistic cases there is simply no reason to *trust* our intuitions. After all, unless you believe (implausibly) that our moral intuitions have a direct pipeline to the truth, you have to worry about the reliability of those intuitions; and the more outlandish the case—the more it falls outside the range of normal human experience—the more reason we have to doubt what our intuitions tell us about that case. So even if there *are* cases where, intuitively, consequentialism seems to permit too much, there is no reason to give our intuitions about those cases very much weight.

(More radically still, some consequentialists simply insist that intuitions about specific cases should carry little or no weight *in general* in assessing moral theories. On this view, even if there is a completely realistic version of the organ case, and even if it is conceded that consequentialism endorses an act—killing Chuck—that seems quite unacceptable intuitively, this will still have little or no force as an objection to consequentialism.)

These, then, are some of the ways in which a consequentialist can respond to a proposed counterexample. He can concede that it would be unacceptable for a moral theory to endorse a given act but insist that consequentialism doesn't actually endorse that act; he can admit that consequentialism endorses the act in question but insist that the given act is not actually intuitively unacceptable (given the circumstances); or he can admit that consequentialism endorses an act that appears intuitively unacceptable but argue that intuition is not—in this case, at the very least—to be trusted. In actual practice, of course, consequentialists typically avail themselves of all of the various replies that I have been sketching, using different answers in response to different specific cases. In this way, they hope to disarm the array of intuitive counterexamples.

How successful is the consequentialist strategy? This is, of course, a matter of some controversy; and like the others considered in this book, I won't try to settle it here. Many find the various consequentialist maneuvers—whether singly, or in combination—sufficiently persuasive so that consequentialism remains a plausible theory of normative ethics. (The discussion from 1.3 should also be recalled: it may be that no theory fits all our intuitions perfectly; so if consequentialism is sufficiently attractive on other grounds, it may be that we are justified in accepting it despite its having counterintuitive implications in certain cases.) Others, however, find the intuitive objections to consequentialism sufficiently compelling that they are forced to conclude that it is inadequate as a normative theory. And if this is the conclusion that one accepts, then of course it is not difficult to find a plausible diagnosis of the source of consequentialism's flaws: consequentialism will seem implausible because of its insistence that goodness of outcomes is the only morally relevant normative factor. A more adequate moral theory will need to recognize the existence of constraints as well; it will need to be deontological.

But for those who do want to move beyond consequentialism, the question arises: what further normative factors are to be recognized? Note that it won't provide much guidance simply to say that we want to move to a deontological system. In principle, after all, deontological theories can differ considerably as to the precise constraints that they accept; so to say that a theory should be deontological just doesn't go very far toward specifying its content. But if the goal is, rather, that of trying to capture something like *commonsense morality* along deontological lines, then we are much further along. We can use the familiar rules of commonsense morality to identify further normative factors that intuitively seem to possess intrinsic moral significance.

In point of fact, we have already identified one further factor that, it seems, must be embraced if we hope to capture something like our ordinary moral intuitions: if an act involves *doing harm*, then this is a highly relevant fact about it, and it weighs in heavily against the moral permissibility of the act. Apparently, then, any normative theory that hopes to capture our ordinary moral intuitions will have to incorporate a constraint against doing harm. Obviously, this is not to say that, with this addition, the list of normative factors is now complete. There may well be other factors that need to be accepted as well—other constraints that must be recognized. The point is simply this: if we are going to move beyond consequentialism, then—whatever other factors we may want to add as well—at the very least we will want to add a constraint against doing harm.

(A final comment about terminology may be in order. Strictly speaking, normative factors are one thing, and the constraints based on them—if any—are another. After all, as we have already noted, a theory might recognize an extra normative factor—such as harm doing—without investing it with the weight and significance necessary to ground a constraint. And from a logical point of view, at any rate, even if a theory does incorporate constraints, it still seems appropriate to distinguish between the normative factor *per se*—which is the relevant feature that a given act might possess or lack—and the *constraint*, which is the prohibition against performing acts that possess the feature in question. Nonetheless, it doesn't seem seriously misleading to sometimes speak as though the constraint itself were the extra normative factor, or to say that a relevant factor in determining the status of an act is that it violates a given constraint. And so, unless context requires otherwise, I will often speak interchangeably of the constraint-generating normative factors and the constraints that these factors generate.)

3.2 Thresholds

As we have seen, many people believe on reflection that harm doing is a factor of intrinsic normative significance; the fact that an act involves doing harm provides a reason not to perform that act. But how much weight

should this factor have? If in some case doing harm is the only way to produce the best results overall, which factor is more important—producing good results, or not doing harm?

It may seem as though we have already answered this question. After all, I have suggested that virtually everyone who accepts the intrinsic normative significance of harm doing gives this factor sufficient weight to override goodness of results in cases like that of the organ transplant. That is to say, they accept a *constraint* against doing harm: it is morally forbidden to kill one innocent person (for example, chopping up Chuck)—even if this is the only way to produce an even better result (saving four more lives overall). But there is still room for disagreement, for the question can be raised as to whether or not this constraint is absolute, or whether it too can be overridden when enough is at stake.

Suppose, for example, that killing an innocent person was necessary not merely to save five lives (as in the organ transplant case)—but a *hundred*, or a *thousand*, or a *million!* (Imagine that killing the one is the only way to prevent a nuclear war; or that the innocent person's body must be used to produce a serum that will provide a cure for a plague that will otherwise kill millions.) Is it still morally impermissible to kill the one innocent person—even with *so much* at stake? Even if this is the only way to prevent a catastrophe?

Some people find themselves still inclined to embrace the constraint, even in the face of such catastrophes. They are *absolutists* about the constraint against doing harm: harming an innocent person is morally forbidden, no matter how horrible the results will be otherwise, no matter how much good could be done. The importance of avoiding harm doing *always* outweighs the importance of producing good results (even when the “good results” include the prevention of bad results.)

Many other people, however, find this absolutist attitude toward the constraint unacceptable. They believe that the constraint against doing harm can itself be outweighed, if *enough* is at stake. Presumably, killing an innocent person is morally forbidden even if this is the only way to save five, ten, or maybe even a hundred or a thousand people—but at *some point*, when the amount of good that needs to be done is great enough, the constraint is overridden, and it is morally permissible to act.

Those who reject the absolutist attitude toward the constraint against harming are *moderate* deontologists. They believe that the constraint has a *threshold*: up to a certain point—the threshold point—it is forbidden to kill or harm an innocent person, even if greater good could be achieved by doing it; but if *enough* good is at stake—if the threshold has been reached or passed—then the constraint is no longer in force, and it is permissible to harm the person.

Should we say that in such a case it is permissible to *violate* the constraint? This way of putting it seems needlessly paradoxical, since talk of

“violating” a rule normally suggests that one has acted impermissibly, whereas the idea here is that the act in question is in fact perfectly permissible, given the pressing circumstances. But we could say, perhaps, that the act *infringes* the constraint—since it is undeniably the kind of act that the constraint normally forbids. If we adopt this convention, then we can say that moderate deontologists believe that not all constraint infringements are forbidden. Normally, of course, it does violate the constraint to infringe it; but when enough is at stake—when the threshold has been met—infringing the constraint is morally justified.

Consequentialists sometimes argue that the attempt to accept deontology while rejecting absolutism is incoherent. After all—they claim—if you are prepared to infringe the constraint on the basis of the good that can be achieved by doing so, then you have in effect admitted that consequences are really the bottom line; and so, whether you realize it or not, you must be some kind of a consequentialist. Interestingly, this is a charge that is often seconded by deontological absolutists, who argue that if you recognize the intrinsic significance of the constraint against doing harm, then you cannot coherently be prepared to abandon or disregard that constraint merely because of the good that this might achieve. Sometimes, then, both sides insist that the only consistent positions are either consequentialism or deontological absolutism; moderate deontology, it is said, is a confused or incoherent position.

I think, however, that such arguments are mistaken. Moderate deontologists can plausibly insist that their position does not reveal itself to be “at bottom” consequentialist, even though they acknowledge that constraints can sometimes be outweighed by good consequences. Consequentialists, after all, believe that goodness of results is the *only* factor with intrinsic significance; it is, therefore, *always* permissible to perform the act that will lead to the best results. But moderate deontologists are *pluralists* who believe in the intrinsic significance of harm doing as well; and normally—that is, when the threshold for the constraint has not been met—it will be morally forbidden to harm someone, even if this will indeed bring about better results overall. Moderate deontology is thus a genuine alternative to consequentialism.

Nor is there anything necessarily incoherent in recognizing the existence of constraints but insisting that these constraints can, for all that, sometimes be outweighed by goodness of results. This is, after all, just what we might expect from a pluralistic view, one that accords intrinsic significance to both goodness of results *and* harm doing. For if there are two (or more) factors, then it shouldn't be surprising if in some cases the first factor outweighs the second, while in other cases the second outweighs the first. Since harm doing has a significant amount of weight in its own right, it will often outweigh goodness of results: this yields the constraint against harming. But as the amount at stake in any given case increases, the importance of this second factor will presumably increase as well—and this might well

yield a *threshold* for that constraint. There is nothing at all incoherent in the suggestion that—despite the intrinsic significance of harm-doing—in some cases goodness of results can be great enough to outweigh the importance of avoiding doing harm, so that, on balance, it will be permissible to infringe the constraint. Moderate deontology is thus a perfectly coherent form of deontology. (But this is not to say that there is anything necessarily incoherent about absolutist deontology either: even if goodness of results has intrinsic normative significance, it might still be that it is *always* outweighed by the importance of not doing harm.)

For those who are attracted to moderate deontology, the question arises of where, exactly, the threshold for the constraint against harming should be located. How *much* must be at stake before it becomes permissible to infringe the constraint? In principle, it should be noted, the range of possible answers is tremendous. Of course, if the constraint is to have any force at all, the threshold must be set above zero. And if absolutism is to be avoided, the threshold must be set at a finite amount. Still, this obviously leaves a huge range of possibilities (an infinite range, in fact). A low threshold might permit killing one to save ten. (Indeed, an extremely low threshold might even permit killing one to save *two*!) A high threshold, in contrast, might only permit killing one to save millions, or billions.

Wherever the line is drawn, the moderate deontologist faces a further challenge (it can be raised by both consequentialists and deontological absolutists): why draw the line *there*? Since the threshold could always have been placed higher or lower, any specific location is bound to seem rather arbitrary. How can it possibly be defended? (By contrast, the consequentialist and the deontological absolutist have staked out the two end points of the continuum; whatever their other drawbacks, the two extremes at least seem less arbitrary.)

Most moderate deontologists, however, will not be greatly distressed by this objection. They will admit that, in some sense, any location for the threshold will seem arbitrary, since it could always have been higher or lower; but they will note that such an appearance of arbitrariness will be present *whenever* a theory needs to trade off two factors against each other. So long as there are two or more factors that need to be balanced, a trade-off schedule of some sort will be required. And in principle, the question can always be raised as to why one particular schedule should be accepted rather than another. But to ask this is to raise an epistemological question: what evidence supports a given schedule as opposed to some alternative? And the answer, presumably, might turn on any number of considerations, including, for example, the extent to which a given schedule fits our intuitions. Since there is no particular reason to think that such evidence cannot be marshaled to help locate the threshold for the constraint against doing harm, moderate deontology is—on the face of it—no worse off than any other form of pluralism.

Presumably, the precise location of the threshold is something that different moderate deontologists will differ over. But it does seem safe to say that if the moderate deontologist hopes to capture something like commonsense morality, then—whatever its exact location—the threshold will have to be fairly high. (It is, however, rather obscure as to exactly *how* high the threshold would have to be to capture our ordinary moral intuitions. It is also unclear how much *precision* is appropriate in specifying the threshold: should we, for example, expect that the threshold can be captured in a precise number, say, “exactly 100 lives”? Or will the threshold be fixed only roughly and imprecisely, “somewhere around 100” or “somewhere in the range of 75 to 125”? These are important issues; but I am going to put them aside.)

Asking for “the” location of the threshold against doing harm may create the impression that there is some single, fixed amount that must be at stake before the constraint can be infringed. On reflection, however, it is clear that this is rather implausible. Far more likely is that the level of the threshold depends on the size of the harm that needs to be done to bring about the good results. For it is important to note that harms do indeed come in different sizes and different kinds; and it is implausible to think that the same amount must be at stake, regardless of the particular harm involved. For example, as we have seen, many people have the intuition that one cannot *kill* an innocent person unless some rather large number of lives can be saved. But presumably most people believe it would take far less to justify inflicting some *minor* harm. It might well be permissible, say, to punch an innocent person in the nose, even if this would save “only” *one* life! Similarly, it seems as though a great deal must be at stake before it would be permissible to cause someone excruciating physical pain; but inflicting some mild discomfort would presumably involve a far lower threshold.

Accordingly, the moderate deontologist might well want to suggest that the level of the threshold in any given case is a function of the nature of the harm involved. On the most simple view of this sort, the level of the threshold is determined by multiplying the *size* of the harm by some fixed amount (so the level of the threshold will vary, but the “multiplier” will be constant). Obviously, however, more complicated functions are possible as well. In addition, the level of the threshold might well be sensitive not only to the size of the harm but also to the *kind* of harm involved. For example, perhaps the threshold is lower if the act merely involves damage to someone’s property, rather than to their body. It is far from obvious what a plausible “threshold function” would look like; and it must be admitted, I think, that moderate deontologists have done little toward filling in the details.

There is a second way in which the need for thresholds for the constraint against harming can arise. Consider the fact that many of our actions involve the *risk* of harming someone else. Driving a car, for example, clearly runs some chance of killing a pedestrian; less obviously, so does riding a bi-

cycle. Few acts, if any, are altogether risk free, although, obviously enough, the level of risk can vary tremendously from act to act. It is also true that few acts, if any, are absolutely guaranteed to cause harm; usually there is at least some chance, however small, that things will end up all right. Now what is a deontologist—any deontologist, absolutist or moderate—to say about these facts? How does the constraint against doing harm apply to cases that only involve the *risk* of doing harm?

One's answer here might well depend on whether one accepts an objective or a subjective account of rightness (see 2.6). An objective deontologist might choose to say that issues involving risk—the mere *chance* of doing harm—are simply irrelevant to the question of whether one has *in fact* violated morality. On this approach, you haven't violated the constraint unless you actually *do* harm. The mere fact that you ran the *risk* of violating the constraint doesn't mean that you actually violated it. (Of course, if the risk is high, you might well be morally blameworthy for taking this chance; but such an evaluation of the *person* shouldn't be confused with the distinct question of whether the act itself violates morality.)

Those attracted to subjective accounts of rightness are likely to prefer a different approach. Consider an act that has a very high risk of causing harm, but imagine that luckily—unexpectedly, and unforeseeably—the harm isn't actually produced. According to the objective deontologist, the agent hasn't actually done anything wrong: since no harm was done, the constraint wasn't violated. But the subjective deontologist finds this an implausible claim: whether the act is right or wrong cannot be a matter of whether or not the harm *happens* to occur; rather, it is a matter of what the agent *believes* (or what it would be reasonable to believe). On this approach, to knowingly perform an action that runs a high risk of causing harm is to act immorally. Thus, the constraint against harming must be understood as a constraint against imposing the *risk* of harm.

But this raises the question: how *great* a risk? That is, what *level* of risk of harm is ruled out by the constraint? It won't do to suggest that the constraint rules out imposing any kind of risk of harm at all: for as we have already noted, almost every act carries *some* risk of harming others. Nor does it seem plausible to claim that the constraint rules out only "high" or "significant" levels of risk. Aside from the vagueness of such labels, there is the deeper problem that risks can be arranged in a smooth continuum from low to high; and it is hard to believe that the constraint might oppose all acts at a given level of risk while remaining altogether indifferent to acts that are only slightly less risky.

This is the point at which an appeal to thresholds might again prove useful. We have already noted the possibility that the level of the threshold is, at least in part, a function of the size of the harm. Here we can note the further possibility that the level of the threshold might also be (in part) a function of the level of the *risk*. Perhaps acts that run a high risk of causing harm have a

high threshold, while acts that run a low risk have a low threshold: other things being equal, the higher the risk, the higher the threshold.

An approach along these lines would explain why it is normally permissible to perform acts that carry only a slight risk of harming others: the risk is low, and so the threshold is low and easily met; such acts need only produce a little good in order to be justified. But acts with significant levels of risk will normally be forbidden: here the threshold will be high, and so it will only rarely be met. Such an approach avoids the need to introduce sharp and arbitrary discontinuities into the constraint. Running any kind of risk at all *infringes* the constraint; but since the level of the threshold varies with the level of risk, many ordinary acts will not violate the constraint.

Of course, this entire approach is only available to moderate deontologists. Absolutists reject the existence of thresholds for the constraint, and so must find some other solution to the problems of risk. They must either embrace the objective approach (according to which risk per se is irrelevant to the constraint), or—if they accept a subjective account of rightness—they must propose a principled and nonarbitrary way to distinguish between unacceptably high levels of risk and ordinary, acceptable levels.

3.3 The Scope of the Constraint

Despite their differences, moderate and absolutist deontologists are alike in believing that harm doing is a factor with intrinsic normative significance. It is only if this factor has normative weight in its own right that there can be any kind of constraint against doing harm (regardless of whether that constraint is absolute or not). That is to say, there can be a constraint against doing harm only because the very fact that an act involves the doing of harm weighs heavily against performing it.

Up to this point, however, we have done little by way of describing the content of this factor. To be sure, we have considered alternative views about the *weight* that this factor might have, but we haven't said very much about the nature of the factor itself. Given our previous purposes, we were able to get by through the use of examples that drew upon our intuitive understanding of harm doing. But it is time to examine the nature of this factor directly. What exactly is it that constitutes doing harm? Just what does the constraint against doing harm oppose?

A natural proposal is this: we harm someone when we act in such a way as to affect their interests adversely. That is to say, to harm someone is to bring about a lowering of their well-being relative to the level that they would have been at, but for our action. If someone is worse off as a result of our act, then we have harmed that person.

It is important to realize just how much gets classified as doing harm under this proposal. In ordinary usage, talk of someone's being "harmed" often suggests pain, physical injury, or death. And this might lead us to expect

that “doing harm” to someone will necessarily involve *causing* them pain, injury, or death. But in fact, if we accept the proposal to construe harm in terms of acting in such a way as to leave a person less well off than they would have been otherwise, then doing harm will include far more than this. First of all, we need not actually *produce* anything intrinsically bad. It will also count as doing harm if we *interfere* with some process that would have *eliminated* some of these intrinsic bads. Thus, although giving someone a painful disease is certainly one way to harm them, it is equally true that you harm someone if you prevent her from getting cured. Furthermore, the relevant change in well-being need not actually involve any intrinsic bads at all: it will suffice if your act leaves the person with fewer intrinsic *goods*. If you “rob” someone of some of the pleasures that she would have had, but for your act, then this too is a way of harming her. Finally, it may be worth recalling that in most cases all we can directly alter is the distribution of *instrumental* goods or bads; but this too can constitute doing harm. We can certainly harm someone by directly causing her pain; but we can also harm her, indirectly, by destroying her property or taking her money.

It seems, then, that we can do harm to someone in all these ways: by adding something evil or bad to their life, or by removing something good; by interfering with the addition of something valuable, or by interfering with the removal of something bad. What’s more, the value or disvalue of the goods or bads in question can be either intrinsic or instrumental. But even having said all this, it still may not be appreciated just how much is going to be counted as doing harm. If it is true that we have done harm to someone whenever our act has adversely affected their welfare, then apparently it doesn’t matter what the particular causal connection is between our act and the change in their welfare; so long as they would have been better off but for our act, we have harmed them. And it also doesn’t matter how *much* worse off the person is; so long as they would have been better off—no matter how slightly—we have done them harm and violated the constraint.

By this point many deontologists might begin to worry that the constraint against doing harm is being construed too broadly. Presumably it is uncontroversial that a constraint against doing harm will rule out killing someone, chopping off their leg or finger, or even punching them in the nose. But do we really want to say that the constraint is violated whenever we leave someone slightly sadder than he might otherwise have been? Do we really want to say that the constraint is violated whenever we damage a blade of our neighbor’s grass? Or whenever the government imposes a tax, however minimal?

Uncomfortable with these results, the deontologist has various options open. Moderate deontologists can appeal, once again, to the existence of thresholds for the constraint against harming. Perhaps we should admit that the constraint truly is *infringed* in all the cases just mentioned, but insist that in many cases of this sort, at any rate, the constraint is not *vio-*

lated, because the relevant threshold has been met. (If the harm is small, little good needs to be accomplished before the act is justified; and if no good at all is going to be promoted, then perhaps it is indeed immoral to harm another, no matter how slightly.)

Deontological absolutists must take a different tack. One possibility, of course, is to embrace the implications of this “broad” reading of the constraint against doing harm. Perhaps we should admit that the constraint is indeed violated, even in cases involving such minimal harms. If this shows just how difficult it can be to live without violating morality, so be it.

Alternatively, the deontologist might try to find a more “narrow” reading of the constraint against doing harm—so that it rules out fewer types of acts. The difficulty, of course, is finding a plausible way of restricting the scope of the constraint. On some views, for example, only the use of *force* is ruled out by the constraint. But this seems too narrow to capture the full range of cases that the deontologist is likely to want to forbid. (If Captain Hook leaves a poisoned cake for the Lost Boys, and they find it, eat it, and die, this presumably violates the constraint against doing harm—even though no force has been used at all.) Other views interpret the constraint in such a way as to rule out only certain *kinds* of harms. (For example, causing death, pain, or physical injury would presumably be forbidden; but causing mere sorrow, inconvenience, or offense might not be.) But here too there is the danger that in at least some cases the deontologist will find the restricted constraint too narrow. And, at any rate, the deontologist may be hard pressed to offer a convincing explanation as to why the various “excluded” types of harm “don’t count,” as far as the constraint is concerned. Similar difficulties face other approaches. Clearly, none of this shows that a plausible narrow reading of the constraint cannot be provided; but it is, I think, far from obvious what it would look like.

A different question about the scope of the constraint arises if we consider the following case: suppose that George is trapped beneath a fallen tree, and will die unless we free him by cutting off his trapped leg. If we amputate, have we harmed him? And if so, have we violated the constraint?

We get two different answers to the first question, depending on how we interpret the proposal that we harm someone if we act in such a way as to leave him worse off than he would have been otherwise. Is the relevant test whether the person is worse off *overall* than he would have been otherwise? Or is it, rather, whether the person is worse off *in some regard* than he would have been otherwise?

On the first interpretation, we consider only the bottom line: we disregard the specific changes in the various components of well-being and simply ask whether *on balance* the person is worse off than he would have been otherwise. On such a *global* interpretation of doing harm, if we save George’s life by amputating his leg, we have not harmed him. It is true, of course, that we have cut off his leg—but presumably it is better for him to

lose a leg than to lose his life, which is what would have happened if we hadn't amputated. Overall, then, our act does not leave George worse off than he would have been otherwise; so we haven't harmed him.

On the second interpretation, we're not actually concerned with the bottom line at all—that is, where the person ends up *on balance*. Rather, we look at the various individual elements that contribute to changes in well-being (whether intrinsically or instrumentally), and we ask whether the person is worse off with respect to one or more of *them* than he would have been otherwise. On such a *local* interpretation of doing harm, if we cut off George's leg, we have harmed him. It is true, of course, that doing this saves his life and leaves him better off overall. But it is also true that the lack of a leg will adversely affect his future well-being; and so, in cutting off George's leg, we have harmed him.

Ordinary language, I think, finds itself somewhat torn between the local and the global interpretations of doing harm. On the one hand, it seems plausible to say, "I didn't *harm* George at all—I saved his life!" And this seems to support the global interpretation. On the other hand, it also seems plausible to say, "I wish I could have saved his life without harming him at all, but unfortunately I had to cut off his leg." And this seems to support the local interpretation. I suspect that the local interpretation fits common usage somewhat better than the global, but rather than try to resolve this question, let's consider the implications of each for the constraint against harming as applied to the case of George. If you do cut off George's leg, does this violate the constraint?

On the global approach, of course, it's obvious that the constraint hasn't been violated, since you haven't harmed George at all. Thus the only relevant factor is the good that can be done; and this justifies your amputating George's leg, since this is the only way to save his life. And the claim that amputation is justified is, I presume, a plausible one: intuitively, that is, it does seem permissible to act in this case. On the local approach, however, things are less clear. Cutting off George's leg harms him, and so this appears to violate the constraint against doing harm. But how then can we justify proceeding?

One possible answer is that the constraint should be interpreted in such a way that it only rules out doing "global" harm, that is, harming someone overall. After all, even if the local interpretation of "doing harm" better captures common usage, it still might be the case that the *constraint* against doing harm is concerned solely with harm that leaves the person worse off overall. If we adopted this approach, we could admit that cutting off George's leg harms him, yet still insist that there is no violation of the constraint, given that George is better off overall having his leg amputated. Yet even if this approach gives the right answer in this case, it remains rather puzzling as to why the constraint *should* be restricted in this way. If "local" harms are true and genuine harms, and if there is something especially sig-

nificant about the doing of harm, then why should the constraint apply *only* in cases of “global” harm?

Moderate deontologists might try a different approach, appealing once again to the existence of thresholds. Perhaps we should admit that cutting off George’s leg harms him, and thus falls under the scope of the constraint, but insist that this merely *infringes* the constraint—rather than violating it—since enough is at stake to justify the infringement. After all, doing harm in this case saves a life, so perhaps this is enough to meet the constraint’s threshold. If so, then despite the infringement, it will be permissible to proceed.

On reflection, however, the moderate deontologist is unlikely to feel comfortable with the suggestion that the threshold has been met. For consider a different case, where cutting off someone’s leg enables us to save someone *else’s* life! (Suppose that if we cut off Harold’s leg we can use it to make enough serum to save Maude’s life.) Presumably the moderate deontologist believes that normally it is *impermissible* to cut off a person’s leg—even if this is the only way to save someone else’s life. So the moderate deontologist must hold that the constraint’s threshold has not been met in such a case: saving a life is not *enough* of a good to override the constraint against cutting off a leg. But if this is so, then we still lack an explanation of what makes it permissible to proceed in the case of George.

There is, of course, another factor that seems highly relevant, one which we’ve yet to mention. It is George’s *own life* that is saved if we cut off his leg, not that of some third party. Intuitively, it seems to matter whether the beneficiary is the very same person who is to be harmed (as with George) or some other person (as with Harold and Maude). But how, exactly, does this make a difference? One possibility is that the threshold is *lower* when the person being harmed is himself the beneficiary. This might allow us to cut off George’s leg for his own sake, without permitting us to cut off his leg for the sake of someone else.

A different possibility is this. If George will die unless we amputate his leg, then presumably he will approve of our doing so. That is to say, he will give us his *permission*; he will *consent* to our action. Perhaps, then, it is the presence of George’s consent that makes acting permissible. We might think of it like this: the constraint against harming protects the individual against being mistreated in various ways. It provides a moral safeguard against his being harmed *against his will*. But if the individual chooses to forego this protection—if for various reasons he *agrees* to being harmed—then the harm no longer constitutes *mistreatment*. In such cases, the protection that is normally in place disappears, and the constraint loses its force. Thus, given that George consents to our cutting off his leg, our act may well infringe the constraint, but it does not violate it.

A suggestion along these lines seems to accord fairly well with a wide range of ordinary moral judgments. For example, normally it would violate

the constraint for you to punch me in the nose. But if we've agreed to a boxing match, then each of us has given the other permission, and your punching me seems permissible. (Note, incidentally, that I've only consented to *your* hitting me; if Martha comes along while we're sparring and hits me, this still violates the constraint. The constraint only loses its force against those who have my *permission*.) Similarly, normally it would violate the constraint for you to cut one of my rare, prize-winning orchids; but not if you ask me first, and I agree.

This view is not, however, without its own difficulties. If the constraint loses its force when the individual consents to our harming him, then what if the person consents for silly or stupid reasons? What if the person is insane, or irrational? What if he gives permission only because of some further belief which is—unbeknownst to him—simply false? Does the constraint lose its force in all these cases too? (Suppose Martha foolishly believes that if we cut off her toe she'll grow another one, and so consents to our taking it; is the constraint really without force in this case?)

Some of those sympathetic to the thesis that consent eliminates the force of the constraint are prepared to embrace this claim even when the person agrees for poor reasons, mistaken reasons, or no reasons at all. Others, however, attempt to narrow the scope of the thesis. Perhaps consent has its normal effect only when the person is rational, and only if the decision is informed by full and accurate knowledge of the facts. (In many institutional contexts there is a corresponding requirement that one get "informed consent.") Perhaps consent only has its normal effect if it is backed by good and sufficient reasons.

There is a further problem with the suggestion that it is the presence of George's consent that explains why it is permissible to amputate his leg. Suppose that George is trapped by the tree, as before, and that he will die unless we cut off his leg. But imagine now that George is unconscious and cannot be asked for his consent. Imagine, as well, that if we try to bring him to consciousness first, so as to obtain his consent, too much precious time will be lost, and it will be too late to save him. Can't we amputate his leg anyway—since this is the only way to save George's life—even though we're unable to get his permission first? But if this is permissible, then how is it to be justified?

Those committed to the belief that it is consent that is the key to explaining why the constraint lacks its normal force, may now have to *expand* their view. Since there is no *actual* consent in this case, they might appeal instead to a notion of *hypothetical consent*. Presumably George *would* agree to the amputation, if only he knew the facts and were able to decide for himself. Perhaps such hypothetical consent suffices to eliminate the force of the constraint. (It may be worth noting, if only in passing, how an appeal to mutual hypothetical consent might help explain why it is permissible for us to engage in the normal, everyday activities that subject others

to low levels of risk: presumably it would be rational for us to agree to being subjected to such risks by others, in exchange for their permission to engage in such minimally risky actions ourselves.)

By this point, however, those critical of the relevance of consent might begin to wonder whether *consent* is still doing any real work. After all, we've already seen that the presence of consent may not be sufficient; it may need to be backed by good and sufficient reasons. And now we've seen that consent may not actually be necessary either. For despite its name, hypothetical consent is not really a form of consent at all: it simply points to the fact that it *would* be rational to consent, because there are good reasons for agreeing. But if consent only works in the presence of good reasons, and if good reasons are sufficient, even in the absence of consent—why not just drop the reference to consent altogether? Why not just say that the force of the constraint is eliminated by the various facts themselves—whatever they are—that would make it rational to agree? The consent itself would just drop out of the picture.

Of course, if we do take this more critical line, then we still need an account of exactly what sorts of facts suffice to eliminate the force of the constraint, and why. If consent is not the relevant factor, just what is? Furthermore, there may be cases where it is less easy to dismiss the relevance of consent: suppose that George is conscious but *refuses* to consent to our amputating his leg, arguing that he would rather be dead than a cripple. Many people will be less confident that we can still proceed in this case, even though everything is the same—everything, that is, *except* for the fact that George is withholding his permission. This seems to support the view that it is indeed the presence or absence of *consent* that is relevant to whether or not the constraint retains its normal force. Indeed, even those who think that it is still permissible to proceed, despite George's protests, will likely feel that the constraint has been *overridden* in this case, and not that its force has simply been eliminated; if the constraint is to literally lose its force, consent itself still seems necessary. (This last case may also cause problems for the global interpretation of doing harm: if—as the global interpretation insists—our cutting off George's leg to save his life doesn't actually harm him at all, why would it even be relevant whether or not he opposes our action?)

Let me turn now to one final question about the scope of the constraint against doing harm. Suppose that Agnes immorally attacks Victor, without provocation or other justification. Indeed, Agnes is about to kill Victor. Luckily, however, Victor can protect himself and avoid being harmed—but only if he harms, and indeed kills, Agnes first. Is it permissible for Victor to do this? Or would it violate the constraint against doing harm?

Most people think the answer is fairly obvious: it is morally permissible to kill in self-defense. This point is worth stressing. According to this common view, killing in self-defense is morally *justified*. This position therefore

needs to be distinguished from another, according to which killing someone in self-defense may be something we should normally *excuse*, but it is, for all that, *wrong*. After all, it might be argued, a willingness to kill in self-defense is a natural—perhaps even instinctive—human reaction; accordingly, there might not be much point in *blaming* someone who killed another to defend himself. But none of this would show that such acts are in fact morally justified. I take it, however, that the common view is that killing in self-defense is not merely understandable, and perhaps excusable; rather, it is *permissible*. Provided that you have harmed someone only in self-defense, you have not acted immorally at all. (Actually, according to the common view, you need not literally be engaged in an act of self-defense; you're also permitted to defend *others*. For simplicity, however, I'll stick to the case of self-defense.)

Distinguishing, in this way, between these two views draws upon a distinction between excuses and justifications. Circumstances that *justify* are factors that alter the moral status of the act, so that—given all the relevant facts—the act in question is morally permissible. In contrast, circumstances that merely *excuse* don't affect the moral permissibility of the act; they merely alter our willingness to blame the agent should the forbidden act be performed anyway. The common view, I take it, is that self-defense provides a moral justification and not merely an excuse; indeed, since self-defense is justified, there is nothing to be excused.

Nonetheless, although belief in the permissibility of self-defense is widespread, it is not universal. *Pacifists* hold that it is morally forbidden to harm another, even if this is the only way to defend oneself from unjustified attack. After all, they argue, to harm someone else—even a deliberate aggressor—violates the *constraint* against doing harm. It is true, of course, that Agnes is herself bent on violating the constraint; she is about to harm Victor. But the fact that *Agnes* is trying to violate the constraint doesn't make it any less true that *Victor* will be violating the constraint if *he* harms *her* instead. As the old saying goes: two wrongs don't make a right.

Obviously enough, those who reject pacifism will deny that harming in self-defense violates the constraint. They'll claim, instead, that it merely infringes it. But how is this view to be defended?

An appeal to thresholds does not seem especially promising here, since any given act of self-defense may not do all that much more good than harm. If Victor kills Agnes, for example, one life is lost and one life is saved—a result no better than if he had simply let *her* kill *him* instead. So even a relatively low threshold might not be met. Things are improved somewhat, but not much, if we accept the view (considered in 2.5) that evil aggressors may deserve to suffer. On such a retributivist view, the harm to Agnes may actually count as a *good* result, rather than a bad, and so we will be that much closer to meeting the threshold. But as long as the threshold isn't very low, this isn't likely to take us far enough. We are somewhat

further along—perhaps even considerably so—if we take into account the *deterrent effect* that an act of self-defense might have on other would-be aggressors (not to mention Agnes), who may be more hesitant about attacking others if they reasonably fear that they themselves might be harmed in the process. Perhaps all of this together would help us meet a low threshold, or even a moderate one. But, as we have seen, most people accept a rather *high* threshold for the constraint against doing harm; and it seems rather unlikely that *enough* good will be done by normal acts of self-defense to meet such a threshold.

More promising, and probably more intuitively compelling, is the suggestion that the constraint against doing harm simply doesn't *apply* to harming aggressors. On this view, the constraint doesn't actually oppose doing harm *per se*; it only opposes harming the *innocent*. Since Agnes is a deliberate aggressor, the use of harm against her does not fall within the scope of the constraint. Indeed, this view is so widely held that when I introduced the constraint (in 3.1 and 3.2) I often explicitly stipulated that the person being harmed was innocent; I did this because I wanted to make sure that the constraint would seem intuitively relevant.

But why *should* the constraint apply only to the innocent? The existence of pacifists, after all, shows us that it is far from self-evident that harming the guilty is morally permissible. Exactly why is it, then, that Agnes's *guilt* takes her beyond the moral protection normally afforded by the constraint?

Perhaps this is a question that can only be settled by appeal to one or another theory about the *foundations* of normative ethics (something we won't be turning to until 6.1). Maybe the best we can do at this stage is to note that, intuitively at least, most people do find it plausible to suggest that guilt versus innocence is itself a factor with intrinsic normative significance: there are indeed weighty moral reasons that oppose doing harm to someone, but these reasons only exist if the person is innocent; if someone is a deliberate aggressor, these reasons don't apply, or have no force.

Unfortunately, even this answer is inadequate, for it points us toward an overly simplistic account of self-defense. In reality, our beliefs about self-defense are surprisingly complex. For example, even if Agnes is a deliberate aggressor, most people don't believe that the protection normally afforded by the constraint simply disappears altogether. Roughly speaking, Victor must use the *minimal* amount of force necessary to defend himself against the attack. (If Victor knows he can save himself by shooting Agnes in the leg, rather than killing her, that's all he's permitted to do; and if merely brandishing his rifle would suffice to stop her, he can't simply go ahead and shoot her anyway.) Similarly, the harm imposed in self-defense must be *proportional* to the harm that the aggressor is trying to do. (A defender may not be restricted to imposing the *same* size harm as would be imposed by the aggressor, but nonetheless there is a limit: if all that Agnes is trying to do is to punch him in the nose, Victor isn't permitted to *kill* her—not even

if there is no other way to stop her.) Yet if the constraint against doing harm simply didn't *apply* in cases where the person being harmed is a deliberate aggressor, then neither of these requirements would make any sense. Apparently, then, the constraint against doing harm must still be in force—although somehow altered or modified in various ways—even if the person being harmed isn't *innocent*.

What's more, the permissibility of self-defense also seems to extend to some cases where the person being harmed isn't a deliberate aggressor at all! Imagine that Malcolm is pointing a loaded gun directly at you and is about to shoot it—mistakenly believing that it is only a harmless toy. Malcolm is clearly not a deliberate aggressor. Yet many people believe it would still be permissible to harm him, were this the only way to prevent him from shooting the gun. Apparently, then, an appeal to self-defense can also justify harming *innocent threats*, and not only harming the guilty.

Of course, an innocent threat is still a *threat*—however unwittingly—so it is possible that accommodating these intuitions will require only a minor modification of the account of self-defense. Note, however, that the innocence or guilt of the threat seems to make a difference to how *much* harm we are permitted to impose in the course of defending ourselves; so, if anything, the account must grow more complicated still.

And there are still other cases where it seems permissible to harm someone who is not, in point of fact, a threat at all. If the only way to stop Abigail from running you and your family over with her tank is to blow the tank up, this still seems permissible to many people—even if this will also kill the baby that Abigail has wickedly strapped to the front of her tank. Thus *innocent shields* of threats can apparently be harmed as well, in the course of defending oneself.

Obviously enough, by this point we have gone far beyond the simplistic suggestion that self-defense can be accommodated by restricting the scope of the constraint so that it only rules out harming the innocent. An adequate account of self-defense, it seems, will be a rather complicated affair (assuming one can be provided at all). No simple appeal to guilt versus innocence is likely to capture the various subfactors that play a role—intuitively, if nothing else—in determining when one can infringe the constraint against doing harm for the sake of defending oneself. (It should also be noted that this increasingly complex account would grow even more complicated, were we to try to extend it to cover the permissibility of *punishment*. For in punishment, unlike normal cases of self-defense, we harm the guilty party only *after* the fact, that is, when it is too late to prevent the violation from occurring in the first place.)

One final question. If you do believe in the permissibility of self-defense, does this show that you are not an absolutist? This is presumably what pacifists would say, insisting—plausibly—that one can hardly be an absolutist about the constraint against doing harm if one is prepared to make an

exception to that rule in cases of self-defense. Yet the permissibility of self-defense has been accepted by many self-styled absolutists. They would presumably argue that they do not literally accept a constraint against doing harm *per se*, but rather a more complex constraint, which can be roughly (and only roughly) summarized as being a constraint against doing harm to the innocent; and with regard to *this* constraint, they might insist, they *are* absolutists: no exceptions can be made to it (for example, it can't be infringed simply because of the good that might thereby be accomplished). This too seems a plausible response. Yet what then prevents the moderate deontologist, who accepts thresholds, from insisting that she too is an absolutist—an absolutist with regard to the even *more* complex constraint which can be summarized (again, only roughly) as being a constraint against doing harm to the innocent unless enough is at stake?

The moral, I suspect, is this: talk of “absolutism” reveals little, if anything, about a person's normative theory. There is simply no useful substitute for detailed knowledge concerning which normative factors a person accepts, and how those factors are thought to interact.

3.4 Doing and Allowing

Given the questions about scope raised in the previous section, it may be misleading in certain contexts to talk as though the deontologist accepts a simple constraint against “doing harm.” In reality, as we have seen, the situation may be rather more complex than this: roughly speaking, the constraint may only oppose certain *ways* of harming someone; and (again, roughly) if the person isn't innocent, or if she consents to being harmed, the constraint may not oppose doing harm at all. Nonetheless, for many purposes we can put these qualifications aside. After all, in many “paradigm cases” of doing harm (for example, killing an innocent person against their will, where the person is neither a threat, nor a shield of a threat, and so on), these qualifications simply don't come into play. Accordingly—and if only for the sake of simplicity—it doesn't seem seriously misleading if we continue to talk of the deontologist's accepting a constraint against “doing harm.” We can, of course, still bring in the various qualifications as needed.

Now as we noted at the start of this chapter (in 3.1), such a constraint cannot be accepted by the consequentialist. It is worth recalling precisely why this is so: it is due to the fact that—unlike the deontologist—the consequentialist rejects the intrinsic normative significance of *harm doing*. Of course, as we also saw, the mere acceptance of this factor does not, by itself, suffice to ground a *constraint*; that will result only if one goes on to invest this factor with sufficient weight. But the fact of the matter is that the consequentialist denies that harm doing *per se* has any intrinsic moral significance at all.

But how can this be? Surely the consequentialist recognizes that doing harm to someone can make the results worse overall. How then can the