CONFLICTING REASONS

DEREK PARFIT
University of Oxford
All Souls College
derek.parfit@all-souls.ox.ac.uk

ABSTRACT
Sidgwick believed that, when impartial reasons conflict with self-interested reasons, there are no truths about their relative strength. There are such truths, I claim, but these truths are imprecise. Many self-interested reasons are decisively outweighed by conflicting impartial moral reasons. But we often have sufficient self-interested reasons to do what would make things go worse, and we sometimes have sufficient self-interested reasons to act wrongly. If we reject Act Consequentialism, we may have to admit that we sometimes have sufficient or even decisive impartial reasons to act wrongly. But these are early days. We may be able to resolve some of these disagreements.

KEYWORDS
Reasons, ought, moral, impartial, self-interest, how well things go, harming as a means, Rational Egoism, Act Consequentialism, Common Sense Morality, the argument from disagreement, skepticism.

1. SIDGWICK'S PROFOUNDEST PROBLEM

Sidgwick asked:

> How should I live? What should I care about, and what should I do?

Two answers seemed to Sidgwick to be clearly true. There are, he believed, two supremely rational aims: *Our Own Good* and *Universal Good*. As some Rational Egoists claim:

> We have most reason to do whatever would be best for ourselves.

As some Impartial Rationalists claim:

> We have most reason to do whatever would be impartially best, by being, on the whole, best for everyone.

These beliefs might both be true. In what Sidgwick called a 'rational Universe', Our Own Good would never conflict with Universal Good. But Sidgwick reluctantly concluded that the actual Universe is not rational. He believed that

> (A) in some cases, one of our possible acts would be impartially best, but another act would be best for ourselves.

These beliefs together imply that

> (B) in such cases, we would have most reason to act in each of these ways.

But this conclusion is a contradiction, which couldn't possibly be true. We couldn't have more reason, or most reason, to act in each of two different ways.

I claimed that, to avoid this contradiction, Sidgwick revised his beliefs. According to what I called Sidgwick's

> *Dualism of Practical Reason,* or *DPR*: We have most reason to do whatever would make things go impartially best, unless some other act would be best for ourselves. In such cases, we would have sufficient reasons – or enough reason – to act in either way.

Sidgwick made some remarks which suggest DPR. In his brief autobiography, Sidgwick wrote:

> No doubt it was, from the point of view of the Universe, reasonable to prefer the greater good to the lesser, even though the lesser good was the private happiness of the agent. Still it seemed to me undeniably reasonable for the individual to prefer his own. The rationality of self-regard seemed to me as undeniable as the rationality of self-sacrifice. [1]

There is no contradiction here, since it might be reasonable and rational to have either of these conflicting preferences.

As de Lazari-Radek and Singer point out, it is not clear that Sidgwick did revise his beliefs by accepting DPR. Sidgwick never dropped his claim that, if we sometimes have to choose between some act that is impartially best and some act that is best for ourselves, 'Practical Reason' would not be 'consistent with itself'. Such cases would

> force us to admit an ultimate and fundamental contradiction in our apparent intuitions of what is Reasonable in conduct [...] from this admission it would seem

[1] ME xx.

> to follow that the apparently intuitive operation of the Practical Reason, manifested in these contradictory judgments, is after all illusory. [2]

Sharing my high opinion of Sidgwick's *Methods*, de Lazari-Radek and Singer write:

> If the conclusion of the best book ever written on ethics is that some of our apparently most solid and carefully examined intuitions about practical reason are illusory, this poses a serious problem for anyone who, like Parfit, defends the view that we can know some ethical judgments to be objectively true because they are based on reason.

According to what we can call this

> *Objectivist's Problem*: If there are contradictions between some of our most carefully considered normative intuitions, we cannot justifiably believe that we know some objective normative truths.

Sidgwick, I believe, overstates this problem. By revising his beliefs in the way described by DPR, Sidgwick could avoid this contradiction. He could also deny that our need to revise some of our intuitive beliefs shows that no such beliefs could be justified. When Sidgwick calls these beliefs 'intuitive', he is not referring to some mysterious faculty with which we can somehow be in causal contact with normative truths. He means that, as in the case of our logical and mathematical beliefs, we can reach some true normative beliefs, not by observing the world and doing experiments, but merely by thinking carefully about certain questions. When Sidgwick adds that some intuitive beliefs are *self-evident*, he does not mean that these beliefs could not be mistaken. He means only that what justifies these beliefs is not that we can derive them from other true beliefs, or that we have evidence in their favour, but their *content*, or what we are believing. Any such belief, he writes, 'may turn out to have an element of error.' There were such errors, he could claim, in his earlier intuitive beliefs which together implied the contradictory conclusion that we sometimes have most reason to act in each of two different ways.

Though Sidgwick could avoid this contradiction by accepting DPR, he would have found this view deeply disappointing. Sidgwick hoped that, when one act would be impartially best but another act would be best for ourselves, Practical Reason would tell us what to do. If DPR were true, Practical Reason would give us no guidance, being 'divided against itself.'

[2] ME 508.

Sidgwick's problem can be restated in moral terms. It seemed to Sidgwick to be clearly true both that

> we have most reason to do our duty, by doing what would be impartially best,

and that

> we have most reason to do what would be best for ourselves.

These beliefs imply that

> when one act is our duty but another act would be best for ourselves, we would have most reason to act in each of these ways.

But this conclusion is another contradiction, which couldn't possibly be true. As Sidgwick wrote, if duty and self-interest sometimes conflict,

> the whole system of our beliefs as to the intrinsic reasonableness of conduct must fall [...] we may perhaps still find in the non-moral universe an adequate object for the Speculative Reason, capable of being in some sense ultimately understood. But the Cosmos of Duty is thus really reduced to a Chaos: and the prolonged effort of the human intellect to frame a perfect ideal of rational conduct is seen to have been foredoomed to inevitable failure.[3]

These magnificently sombre claims are, I claimed, overstatements. Sidgwick would avoid this contradiction if, as I suggested, he revised his beliefs. According to what I called Sidgwick's

> *Dualism of Duty and Self-Interest*, or *DDS*: When we have a duty to act in one way, but another act would be best for ourselves, we would have sufficient reasons – or enough reason – to act in either way.

There is again no contradiction here. But if DDS were true, that would gravely undermine morality. We would have to admit that

> whenever it was against our interests to do our duty, we would have sufficient reasons to act wrongly.

Sidgwick was not unusual in regarding such cases as undermining morality. Many other people have believed that, if there could be such cases, morality would be undermined. There cannot be such cases, many of these people have believed, because wrong acts would all be punished either in Hell or in some reincarnated life on Earth. Reid wrote that if it could be against our interests to do our duty, we would be faced with 'this miserable dilemma, whether to be

[3] ME First Edition (1874) 473.

a knave or a fool'. We would be a knave if we didn't do our duty, but a fool if we did.

2. MORAL AND SELF-INTERESTED REASONS

De Lazari-Radek and Singer suggest a solution to Sidgwick's problem. Our most fundamental normative beliefs are not beliefs about the natural world, for which we might have empirical evidence. These beliefs are like other non-empirical beliefs, such as logical and mathematical beliefs. We have some of these beliefs because they are clearly or obviously true. When we know, for example, that

(C) some argument is valid and has true premises,

it is clear both that

(D) this argument's conclusion must be true,

and that

(E) we have a decisive reason to believe this conclusion.

Though (E) is a normative belief, this belief is as clearly true as the logical and modal belief stated by (D).

Of the contradictory beliefs that seemed to Sidgwick to be clearly true, one was the Rational Egoist's belief that

(F) we have most reason to do whatever would be best for ourselves.

We can challenge this belief, de Lazari-Radek and Singer suggest, with an evolutionary debunking argument. Most of us care about our own well-being much more than we care about the well-being of strangers. Natural selection explains this fact, since those early humans whose genes made them more self-interested would have been more likely to survive and spread these genes. When our ancestors came to believe that they had most reason to do what would be best for themselves, this belief merely endorsed these self-interested motives. These motives and this belief would have been reproductively advantageous whether or not this belief is true. When we realize that we were caused to believe (F) in a way that was unrelated to whether this belief is true, that casts strong doubt on this belief.

No such argument applies to the Impartialist's belief that

> (G) we have most reason to do whatever would be impartially best, by being, on the whole, best for everyone.

This belief cannot be similarly challenged, because it cannot be given any such evolutionary explanation. Wholly impartial motives and beliefs would have made early humans less likely to survive and spread their genes. Since there is no such debunking explanation for the Impartialist's belief, we can claim that we have this belief because, as rational beings, we have recognized the truth that everyone's well-being matters equally. We can see that our own suffering is not worse than the equal suffering of anyone else, or any other sentient being.

This argument has, I believe, some force. But, as de Lazari-Radek and Singer concede, this argument is not decisive. They claim only that,

> if the rationality of egoism can thus be put in doubt, we can tentatively conclude that all reasons for action are impartial.

I suggested another argument against (F). Rational Egoists believe that we ought to be equally concerned about all the parts of our life. It is irrational, for example, to care more about those pleasures and pains that we shall have in the nearer future. A mere difference in *when* we shall be in pain is not a difference in the badness of this pain. Impartial Rationalists might similarly claim that we ought to be equally concerned about everyone's life, since a mere difference in *who* is in pain is not a difference in the badness of this pain.

Like the evolutionary debunking argument against (F), this argument seems to me to have some force, but is not decisive. The distinction between different times may be relevantly different from the distinction between different people. The separateness of persons has been called 'the fundamental fact for ethics', and this fact, Sidgwick plausibly believed, had great rational significance. Most of us would find it hard to believe that we have no reasons to be more concerned about our own well-being.

If we admit that we have these reasons, we can respond to Sidgwick's problem in another way. Sidgwick seems to have assumed that, when we have impartial reasons to act in one way, and self-interested reasons to act in some other way, these reasons are incomparable, in the sense that there are no truths about their relative strength. This assumption is, I have claimed, mistaken. Some self-interested reasons are decisively outweighed by some conflicting impartial reasons. We would have decisive reasons, for example, to do what would be slightly worse for ourselves if this act would save some stranger's life. Sidgwick's claim should have been only that, because such truths are imprecise, we often have sufficient reasons either to do what would be impartially better or to do what would be better for ourselves.

Similar remarks apply when it would be wrong to do what would be better for ourselves. Some self-interested reasons are decisively outweighed by some conflicting moral reasons. But when some wrong act would be much better for us than any of our other possible acts, we would sometimes have sufficient reasons to act wrongly in this way. These reasons would be different on different moral views. Suppose first that, in

> *Case One*, after some shipwreck you could save either your own life or some stranger's life. Though you could expect to live for many happy years, this stranger could expect to live happily for slightly longer, so that if you saved this stranger's life you would be likely to make the outcome slightly better.

Act Consequentialists would believe that you would be acting wrongly if you saved your own life. But since this act would be likely to make the outcome only slightly worse, you would have only a weak moral reason not to act in this way. As Act Consequentialists should admit, this weak moral reason would not be stronger than your strong self-interested reason to give yourself many more years of happy life. You would have a sufficient reason to act wrongly by saving your own life.

There are many actual cases of this kind. We could often make the outcome slightly better by doing something that would be much worse for ourselves. Act Consequentialists should admit that we would not have decisive reasons to act in these ways. Weak moral reasons do not always outweigh strong self-interested reasons.

Most of us are not Act Consequentialists. We accept some Non-Consequentialist moral view, such as some form of Common Sense Morality. On such views, we are often morally permitted to give greater weight to our own well-being. In *Case One*, for example, you would have no duty to sacrifice your life so that you could give some stranger a slightly greater benefit.

Suppose next that, in

> *Case Two*, you could save your life, but only by imposing harm on some stranger.

On plausible Non-Consequentialist views, you would be morally permitted to save your life with some act that would cause this stranger to have a nasty bruise. It would be wrong, however, to save your life by imposing on this stranger some great harm. In between these extremes there would be some harms which are the greatest harms that you could justifiably impose when that is the only way to save your own life. If you imposed slightly more than

this justifiable amount of harm, your act would be only slightly wrong. Suppose, for example, that you could justifiably save your life by imposing on some stranger N minutes of pain. If that is true, it could not be very wrong for you to save your life by imposing on this stranger N + 1 minutes of pain. Your self-interested reason to give yourself many happy years of life would not be outweighed by your weak moral reason not to impose slightly more than this justifiable amount of harm. You would have a sufficient self-interested reason to save your life by acting wrongly in this way.

As these examples help to show, we sometimes have only weak moral reasons to avoid acting wrongly. We cannot defensibly believe that these weak moral reasons would always outweigh even the strongest self-interested reasons. That would be like Newman's horrific view that sin is infinitely worse than pain, so that, if all mankind suffered extremest agony, that would be less bad than if one venial sin were committed.

I shall now sum up these conclusions. According to

> *Strong Egoistic Rationalists*: We always have decisive reasons to do what is best for ourselves.

According to

> *Strong Moral Rationalists:* We always have decisive reasons to do our duty, however weak this duty is, and however bad this act would be for ourselves.

On Sidgwick's view,

> we never have decisive reasons to do our duty if this act would be bad for ourselves.

Defensible views, I have claimed, are in between these extremes. Neither self-interested nor moral reasons are always decisive. When it is against our interests to do our duty, we often have decisive reasons to do our duty, but that is not always true.

3. OTHER PROBLEMS

Sidgwick's problem is raised by conflicts between self-interested reasons and impartial moral reasons. Since Sidgwick was an Act Consequentialist, he believed that impartial reasons never conflict with moral reasons. If we accept some Non-Consequentialist moral view, we believe that, in some cases, these reasons do conflict. These cases raise some other problems.

When we are deciding what to do, we can ask:

> Q1: What ought I morally to do? Do I have a duty to act in some way, because every other possible act would be wrong?
>
> Q2: What do I have most reason to do? Do I have decisive reasons to act in some way?

We can call these *the Moralist's Question* and *the Rationalist's Question*. Q2 could be restated as

> Q3: Is there something that I ought to do in the decisive-reason-implying sense?

The Rationalist's Question is, I claimed, more fundamental. That can be shown by supposing that we have decisive reasons to act wrongly, or that we have no reason to care whether our acts are wrong. Morality would then have a status that is like that of etiquette, or some code of honour. If we had no reason to do what we ought morally to do, that would undermine morality. There could be no such truth the other way round. If it would be wrong for us to do what we have decisive reasons to do, that would not undermine these decisive reasons.

Suppose next that, as some people believe, we always have decisive reasons to do our duty and to avoid actingly wrongly. If morality has such supreme importance, that would be true only because we had these decisive reasons, so it would still be these reasons that were fundamental.

We can also ask

> Q4: How would it be best for things to go?

One of two events would be better if we all have stronger impartial reasons to want and hope that this event will occur. It would be in this sense better, for example, if more people's lives are saved.

When Sidgwick discusses what we ought to do, he sometimes seems to be using 'ought', not in a moral sense, but in the decisive-reason-implying sense. When he states his principle of Rational Benevolence, Sidgwick writes:

> the good of any one individual is of no more importance [...] than the good of any other [...].

and

> as a rational being I am bound to aim at good generally [...]. I ought not to prefer my own lesser good to the greater good of another. [4]

Sidgwick here claims that he has most reason to aim at good generally. This claim does not imply that, if he aims at his own lesser good, his acts would be wrong in some distinctively moral sense, by being, for example, acts that are blameworthy, or unjustifiable to others, or acts that would give him reason for remorse and give others reasons for indignation. And though Sidgwick's profoundest problem can be stated in moral terms, this problem is about the conflict between impartial and self-interested reasons.

De Lazari-Radek and Singer make some similar remarks. Singer has written:

> Thus, looking at things ethically is a way of transcending our inward-looking concerns, and identifying ourselves with the most objective point of view possible—with, as Sidgwick put it, 'the point of view of the universe'. [5]

> Ethics takes a universal point of view. [6]

> Moral acts, understood as acts that are justifiable from an impartial perspective [...] are rationally required. [7]

In summing up their view, de Lazari-Radek and Singer write:

> when one of two possible acts would make things go impartially better, that is what we have decisive normative reason to do. [8]

These are not distinctively moral claims, and they suggest that we have decisive reasons to do whatever we have the strongest impartial reasons to do.

We can now return to conflicts between moral and impartial reasons. If we are not Act Consequentialists, we believe that it would sometimes be wrong to do what we have the strongest impartial reasons to do. We should ask whether, in such cases, we might have sufficient or decisive impartial reasons to act wrongly.

Here is another way to describe this question. What is distinctive of all Consequentialist views is the central place, or greatest weight, that Consequentialists give to truths about how it would be best for things to go. Such views can

[4] ME 382-3.

[5] Peter Singer, *Practical Ethics,* third edition (Cambridge University Press, 2011) 293.

[6] Ibid., 11.

[7] Peter Singer, ed. *Does Anything Really Matter? Parfit on Objectivity* (Oxford University Press, forthcoming).

[8] P. 199.

take at least two forms. Some Consequentialists make claims about the *moral* importance of how well things go. According to these

> Act Consequentialists: We ought morally to do whatever would make things go best.

Other Consequentialists make claims about the *rational* importance of how well things go. According to these people, whom I shall call

> *Impartial Rationalists*: We have strong reasons to do whatever would make things go best.

As Sidgwick saw, but did not clearly enough state, Rational Egoism is best regarded, not as a moral view, but as an external rival to all moral views. We can make a similar claim about Impartial Rationalism. If we are not Act Consequentialists, we believe that, in some cases, it would be wrong to do what would make things go best. Act Consequentialism is the *internal* rival to our moral view, since these Consequentialists reject some of our moral beliefs. Impartial Rationalists may not reject these beliefs, since these people may agree with us about which acts are right or wrong. But Impartial Rationalism may be an *external* rival to our moral view, since these Rationalists may believe that, in some cases, we would have sufficient or decisive reasons to act wrongly.

According to Strong Impartial Rationalism, we have such decisive reasons whenever wrong acts would make things go best. As my earlier remarks imply, we should reject this view. On a more plausible view, which we can call

> *Weak High Stakes Impartial Rationalism*: When some wrong act would make things go much better than all our other possible acts would do, we would sometimes have sufficient or even decisive impartial reasons to act wrongly in this way.

There are several kinds of act which most of us would believe to be wrong even when these acts would make things go better. We can take, as our example, some of the acts that would impose serious harm on some people as a means of benefiting others. According to what we can call

> *the Means Principle*: It would often be wrong to kill one person as a means of saving the lives of several other people.

In one much discussed case, which we can call

> *Bridge*, some driverless runaway train is moving towards five people, whom it threatens to kill. You could save these people's lives, but only by causing me to fall from some bridge onto the track, where this train would be stopped by hitting and killing me.

Many of us would believe this act to be wrong.

The Means Principle does not imply that it would always be wrong to kill one person as a means of saving others. Such acts would be justified, this principle allows, if these acts were the only way of saving the lives of very many people, such as a thousand or a million people. If we believe that, in some cases, such acts would be justified, we should admit that, in some other similar cases, such acts would be only slightly wrong. On this view, there must be some rough number of people which is the smallest number whose lives we could justifiably or permissibly save by killing one person as a means. Suppose, for example, that this smallest number is 100 people. On this view, if we killed one person as a means of saving only 99 other people, this act would be wrong. But this slight difference in the number of people whom our act would save could not make this act very wrong. We would have only a weak moral reason not to act wrongly in this way. We would have a strong conflicting impartial reason to act in this way, since this act would cause 98 fewer people to be killed. Since these reasons would conflict, we should ask whether either reason would be stronger.

Since this act would be only slightly wrong, our moral reason would not, I believe, be decisive. This reason would, I believe, be outweighed by our impartial reason to cause 98 fewer people to be killed. This would be a case in which we had a decisive impartial reason to act wrongly.

We can next consider an example in which the stakes are higher, and the conflicting reasons might both be strong. We can imagine a state of the world that was roughly like its actual state in 1945. The US President, we can suppose, could choose between two ways of ending the Second World War. The President has true beliefs about what would be the main effects of these two policies. In what we can call

> *the Nuclear Policy*, an atomic bomb would be dropped on one city in Japan, where it would kill about 100,000 civilians, most of them immediately or very soon.

This policy would swiftly end the war. The Japanese Government would soon surrender, because the use of this vastly destructive new weapon would give the generals in the Japanese Government what they would believe to be an honourable way to admit defeat. In what we can call

> *the Conventional Policy*, the US armed forces would instead invade Japan and would win the war by fighting in conventional ways. These methods would involve much bombing and fighting in highly populated

areas, whose unavoidable side-effects would be to kill at least 300,000 civilians before Japan surrendered.

The Nuclear Policy would be, in one way, better than the Conventional Policy since this policy would cause at least 200,000 fewer civilians to be killed. On one widely accepted view, however, there is a great moral difference between two ways in which, when some nation is fighting a just defensive war, this nation's armed forces might kill some innocent civilians. On this view, it is sometimes permissible to use *tactical* bombing, which is aimed to destroy military targets such as arms factories, even if such bombing would predictably have the side-effect of killing some innocent civilians who live nearby. It is wrong, however, to use *terror bombing*, which deliberately kills civilians as a means of persuading some enemy to surrender. This *Terror Bombing Principle* – which is one version of the Means Principle – is accepted by many of those who think about the morality of war.

If we accept this principle, we may believe that it would be wrong for the President to choose the Nuclear Policy. This policy would involve pure terror bombing, since the President's only aim in dropping this atomic bomb would be to kill innocent civilians as a means of persuading the Japanese Government to surrender.

If we accept the Means Principle, we are not absolutists. We believe that we could justifiably kill one person as a means of avoiding killing some large numbers of other people. But this belief, we can suppose, would not apply to our example. If the President chose the Nuclear Policy he would kill many civilians as a means of avoiding killing only about three times as many other civilians. The proportions here are similar to those in *Bridge*, in which you could kill one person as a means of saving five other people. If we accept the Means Principle, we may believe that the President ought to choose the Conventional Policy and invade Japan, since the Nuclear Policy would be very wrong. The President's moral reason not to choose the Nuclear Policy conflicts, however, with a strong impartial reason. If the President chose this policy, he would end this war in a way that would kill 200,000 fewer civilians. Would this impartial reason outweigh this conflicting moral reason?

This question is difficult, because we are comparing reasons of different kinds. It is easier to compare reasons which are all provided by the effects of our acts on different people's well-being. We can ask, for example, whether we could justifiably impose some great burden on one person to save many other people from smaller burdens. The answer may depend not only on the size and number of these burdens, but also on how badly off these people are.

When we compare the reasons that are given by such facts, we can often plausibly believe that some of these reasons outweigh others.

It is harder to compare the President's reasons for and against choosing the Nuclear Policy. The objection to killing people as a means isn't given by the greater badness of such acts for the people who are killed. When people are killed as a means, that is not worse for these people than being killed as a side effect would be. If this act would be wrong, as the Means Principle implies, this act would be wrong because it is in itself morally bad to kill some people as a means of saving the lives of others. We must compare the President's moral reason not to act wrongly in this way with his impartial reason to cause 200,000 fewer people to be killed.

This impartial reason, we may believe, would outweigh this moral reason. On this view, the President would have a decisive reason to act in a way that was very wrong. That conclusion would be disturbing.

We may instead believe that the President's moral reason would be decisive. If we are not Act Consequentialists however, this belief raises another problem. We may think

> We could always reasonably want and hope that things will go in the ways that would be best.
>
> It would often be best if some people acted wrongly.

Therefore

> We could often reasonably want and hope that some people will act wrongly.

I call this the Argument for Moral Ambivalence.

We can next suppose that we are US citizens, and we know that the President will soon choose how to end the Second World War. We can ask what we could reasonably want or hope that the President will do. We believe that, if the President chooses the Nuclear Policy, he would be acting very wrongly. This fact would give us a reason to hope that the President will not choose this policy. We also know that, if the President makes this choice, he would end the war in a way that would cause at least 200,000 fewer civilians to be killed. This fact gives us an impartial reason to hope that the President will act very wrongly, by choosing this policy. In asking what we could reasonably want or hope that the President will do, we must compare the strength of these conflicting reasons.

Our impartial reason would, I believe, be clearly stronger. Some Non-Consequentialists believe that one murder would be less bad than two acci-

dental deaths, since it would be these deaths that we ought to prevent. The argument for this view is, I believe, flawed, since it overlooks the badness of an attempted murder. It is significant, however, that these people accepted this argument's conclusion. These people found it intuitively acceptable that one person's being wrongly killed would be less bad than as few as only two accidental deaths. We cannot plausibly believe that, if the President chooses the Nuclear Policy, the badness of the President's wrong act would be greater than the badness of as many as 200,000 deaths. We could reasonably want and hope that the President will act very wrongly in this way, so that these many fewer people would be killed.

This conclusion would also be disturbing. If we could reasonably want and hope that the government of our nation will act very wrongly, that threatens the status of our moral beliefs.

4. DISAGREEMENTS

If we are Act Consequentialists, we avoid these conclusions. De Lazari-Radek and Singer suggest that, to defend Act Consequentialism, we can appeal to another evolutionary debunking argument. Natural selection can explain why our ancestors had many of the non-impartial motives that were later endorsed by the beliefs of Common Sense Morality, such as our beliefs that we ought to promote the well-being of our close relatives and other members of our group, tribe, or community. These motives and beliefs were reproductively advantageous. Natural selection can also explain how we became *reciprocal altruists*, or *grudgers*, who benefit only those other people who reciprocate by benefiting us in return. Reciprocal altruists also act in ways that spread their genes. Since these moral beliefs would have been reproductively advantageous whether or not they were true, this evolutionary explanation would undermine or at least weaken the justification of these beliefs.

Natural selection cannot, however, explain how we came to have some other widely held moral beliefs. On the Golden Rule, for example, we ought to treat other people as we would want other people to treat us. Most of us would want other people to benefit us whether or not we benefit them. The Golden Rule therefore tells most of us to be, not grudgers, but *suckers*, who benefit everyone, including people whom we know will never benefit us. Nor can natural selection explain how we came to have several other impartial moral beliefs, such as the beliefs of Act Utilitarians. It was not reproductively advantageous to be-

lieve that the suffering of all sentient beings matters, and matters equally. Since these impartial beliefs were not advantageous, they are not challenged by this evolutionary debunking argument.

These claims about this argument are, I believe, plausible. As de Lazari-Radek and Singer point out, however, though this argument challenges some widely held moral beliefs, it cannot by itself show that we ought to become Act Utilitarians. This argument applies to 'many of our common moral judgments', which include 'all kinds of partial moral judgments', and all moral judgments that are based on 'evolved evaluative atitudes'. But there are several widely held Non-Utilitarian moral beliefs that are not in these ways *partial* or based on such evolved attitudes. As de Lazari-Radek and Singer write elsewhere:

> Other principles, including deontological principles, might be equally impartial – for instance, the principle that lying is wrong, whether one is lying to strangers or to members of one's own community.[9]

Other examples are some of the beliefs that are defended by Kantians or Contractualists. Since this evolutionary argument does not apply to these Non-Utilitarian moral beliefs, Act Utilitarians would have to defend their view by giving other arguments.

Return now to the question whether we can justifiably believe that there are some objective, irreducibly normative truths. Many moral skeptics claim that, given the deep disagreements between many people's moral beliefs, we cannot justifiably believe that there are any such moral truths. Sidgwick was greatly disturbed by such disagreements. As I have said, Sidgwick wrote:

> if I find any of my intuitions in direct conflict with an intuition of some other mind, there must be error somewhere: and if I have no more reason to suspect error in the other mind than in my own, reflective comparison between the two intuitions necessarily reduces me [...] to a state of neutrality. [10]

If we have such deep disagreements with other people, and we seem to have no reason to believe that *we* are the people who are much more likely to be getting things right, we cannot justifiably keep our beliefs, except in significantly weaker forms. We may even be led to doubt that any of us might be getting things right, by having true beliefs, since there may be nothing to get right. I called this *the Argument from Disagreement.*

[9] P. 203.
[10] ME 342.

Sidgwick's Act Utilitarian beliefs often conflict with the beliefs of Common Sense Morality. In the longest part of his *Methods*, Sidgwick argues that these disagreements are not deep enough to count strongly against his Utilitarian beliefs. Summing up his claims, Sidgwick writes:

> the Utilitarian argument cannot be fairly judged unless we take fully into account the cumulative force which it derives from the complex character of the coincidence between Utilitarianism and Common Sense.
>
> If systematic reflection upon the morality of Common Sense thus exhibits the Utilitarian principle as that to which Common Sense naturally appeals for that further development of its system which this same reflection shows to be necessary, the proof of Utilitarianism seems as complete as it can be made. [11]

In one passage de Lazari-Radek and Singer suggest that Utilitarians need not discuss the conflicts between their moral beliefs and Common Sense Morality. Act Utilitarians, they write, could

> draw on evolutionary theory, as well as on Sidgwick's normative arguments in order to reject many widely-shared moral intuitions, while retaining the principle of universal benevolence. Although those who make use of reflective equilibrium in normative and applied ethics typically assume that they should try to achieve an equilibrium between a plausible normative theory and most, or at least many, of our commonly accepted moral judgments, there is no need for them to make this assumption.

These claims are, I believe, too strong. De Lazari-Radek and Singer acknowledge the force of the Argument from Disagreement. They suggest that, to defend our moral beliefs, we should try to defend the view that, in ideal conditions, we would reach agreement with all other 'careful thinkers'. Nor is it enough, as they themselves claim, to appeal to an evolutionary debunking argument. And as my last quotations suggest, Sidgwick's normative arguments mainly consist in his many detailed claims about the relations between his Utilitarian principle and Common Sense Morality.[12]

When I earlier discussed this Argument from Disagreement, I defended the *Convergence Claim* that, in ideal conditions, we would nearly all have sufficiently similar normative beliefs. Many actual moral disagreements can be explained in ways that are *untroubling* in the sense that they do not cast doubt on this Convergence Claim. Some of these disagreements depend on disagreements about non-normative facts. Other disagreements are about border-

[11] ME 422.

[12] Unlike Mill and Moore, whose defences of Act Consequentialism take less than a page, Sidgwick's defence take two hundred pages.

line cases, where we would expect some people to have conflicting beliefs. Two examples are disagreements about abortion and euthanasia. These disagreements do not cast doubt on the nearly universally accepted view that it is wrong to kill innocent people. Other disagreements depend on conflicting religious beliefs, or on distorting factors, such as people's assumption that, if certain moral beliefs were widely accepted, that would be good or bad for them.

Some disagreements cannot yet be explained in these and other untroubling ways. But these are early days. As Rawls and Nagel claim, our moral theories 'are primitive, and have grave defects', and 'ethical theory [...] is in its infancy.' [13] In responding to this argument for moral skepticism, we should therefore ask whether we can rationally hope to make progress in ways that make our disagreements less deep. In considering this question, we should try to avoid wishful thinking, by believing what we hope is true.

In *Volume One* of *On What Matters*, I try to show that if three of the main systematic moral theories are revised in ways that seem to be clearly needed, these theories converge. In what will be my *Volume Three*, I try to show that, as Railton and Gibbard suggest, if three of the main meta-ethical theories take their best forms, these theories also converge. But the most worrying disagreements are not between different systematic moral theories or between different meta-ethical theories, but disagreements about what matters, and about which acts are right or wrong. The deepest disagreements of this kind are between Act Consequentialism and some parts of Common Sense Morality We should therefore ask whether these disagreements can be resolved. If this attempt succeeds, that would support the view that there are some objective normative truths, some of which are moral truths. I start this attempt in the last two chapters of my *Volume Three*, and hope to continue this attempt in what would be my *Volume Four.*

[13] John Rawls, *Theory of Justice*, (Harvard University Press 1971) 52; Thomas Nagel, *Other Minds* (Oxford University Press, 1995) 102.