DAVID MCCARTHY

ACTIONS, BELIEFS, AND CONSEQUENCES*

(Received in revised form 7 May 1996)

G. E. Moore urged this distinction.[1] Whether an agent's act is permissible depends, in part, on what the consequences of the act *will* be, and not on what she believes they will be. Whether she is justified in performing the act (or as some writers say, whether her act is subjectively permissible) depends, in part, on what she *believes* the consequences of the act will be, and not on what they will be.

There are many variations on this. For example, justification may have more to do with what it is reasonable for the agent to believe rather than what she actually believes. And there are many different kinds of moral theory in which these kinds of distinctions can be made. On some theories, only the consequences of acts are relevant to their permissibility; on others, things beside consequences are relevant. But the idea behind these variations seems clear. Insofar as the consequences of an agent's act are somehow relevant to moral claims connected with the act, facts about what the consequences of the act will be are relevant to one cluster of moral claims, whereas facts about what the agent believes they will be, or what it is reasonable for her to believe they will be, are relevant to another cluster, and we do well to separate these facts. Let us call this the doctrine of separation of acts from agents.

In some ways this separation is awkward and artificial. Many of our ordinary act-descriptions involve facts from both domains: consider murder. But while it may be neither desirable nor possible to eliminate such act-descriptions from ordinary moral discourse, there may be a theoretical advantage to drawing this sharp distinction between the grounds of justification and the grounds of permissibility.

We talk of whether an agent was justified in performing a particular act when we address issues of responsibility, praise or blame, excuses, and, on some theories, punishment and liability. In these

Philosophical Studies **90:** 57–77, 1998.

contexts, the agent's beliefs about her action, or the beliefs it was reasonable for her to have, seem more relevant than whatever consequences her action turned out to have.

If that is right, what reason is there for moral theory to concern itself at all with permissibility? One reason is that justification is in some ways parasitic on permissibility. At first approximation, an agent is justified in performing a particular act if and only if it would be permissible for her to perform that act if her beliefs, or the beliefs it is reasonable for her to have, were true. A second reason is that an important part of our moral discourse only appears to make sense when permissibility is understood to be independent of agents' beliefs. Suppose that I am about to give a young child what I (with good reason) believe to be a piece of candy, and I say to myself: "It is not the case that you ought not to give him that." You, however, know that it is rat-poison, and you quickly say: "You ought not to give him that!" Surely what you said was true; and doesn't that show that what I said was straightforwardly false? My beliefs, or what it was reasonable for me to believe, or even your beliefs for that matter, seem to have no bearing on the truth of your utterance.

There are several ways in which the separation of acts from agents might be attacked. Many people think that the structure of an agent's intentions, for example, is relevant to the permissibility of her acts. One might then try to argue that what an agent's intentions are depends on what she believes, so her beliefs really are relevant. Or one might try to argue independently of this that it is incoherent to say that one part of her mental life – her intentions – is relevant, while another part – her beliefs – is not. I will ignore these lines of attack, however. Here I want to argue that the phenomenon which appears to best support the separation of acts from agents, the fact that agents with good reason often do not know what the consequences of their actions will be, cannot be made to fit coherently into our moral theory if we make that separation.

1. Consider

S(*witch*) 1 It is getting dark and I am just about to turn on my lights. I recall having read that day that there is substantial empirical evidence that this will impose a one in a billion risk of death due to a freak electrical discharge on my neighbor, Smith. But the risk is trivial, so I flip the switch. Sadly, my act causes an electrical discharge which causes Smith's death.[2]

I was quite clearly justified in flipping the switch. The risk was trivial, and I had no reason to believe that it would cause Smith's death. But was it permissible for me to flip the switch? It seems implausible to deny that it was permissible; after all, I was clearly not in any way to blame for causing Smith's death. But we know what defenders of the separation of acts from agents would say. They would say that my having been justified in flipping the switch does not entail that it was permissible. As Moore said, there are many things which we may be justified in thinking true which are not true.[3] And there is evidence to support the claim that it was in fact impermissible. If you had told me at the time of my flipping the switch that I ought not to cause Smith's death I would not have doubted you for a moment.

I should say that my having imposed a small risk on Smith in S1 is consistent with its having been the case that it was necessary given the set up of the wiring and the laws of nature that my flipping the switch would cause Smith's death. Talk of risk is at least sometimes a reflection of the appropriate epistemic perspective, and does not presuppose that the world is non-deterministic.

Defenders of the separation of acts and agents even have an argument for the claim that it was impermissible for me to flip the switch, the argument from the objectivity of ought.[4] Consider

S2 Same as S1, except as I am about to flip the switch I notice out of the corner of my eye that the wiring is in a very dangerous condition. I now have every reason to believe that if I flip the switch, I will thereby cause Smith's death.

As it happens I had earlier that day read in my horoscope and had thereby come to believe that I could do no harm that day. But that makes no difference: I surely ought not to flip the switch; my bizarre belief has no relevance. We might say: what it is permissible for an agent to do depends on the way the world really is, not on the way she thinks it is. Likewise, it might be argued, my not having had any reason in S1 to believe the wiring was in a dangerous condition has no relevance, and I was mistaken in thinking that it was permissible for me to flip the switch.

A natural response to this argument is: 'Although it is true that what it is permissible for an agent to do does not depend on the way she happens to think the world is, it does depend on the way it is reasonable for her to think the world is. It was reasonable for you in

S1 to believe that it was not the case that if you flip the switch you will thereby cause Smith's death, but unreasonable in S2.'

But consider

S3 Same as S1, except as I am about to flip the switch I think to myself: "It is permissible for you to flip the switch." But along comes Bloggs the electrician. She notices that the wiring is in a very dangerous condition, and says: "You ought not to flip the switch."

It is very plausible to think that Bloggs is speaking truly in S3, and that seems to contradict my thinking to myself at the time: "It is permissible for you to flip the switch." Although my belief about the state of the world was reasonable, the truth of Bloggs' assertion seems to show that I was just wrong in thinking that it was permissible for me to flip the switch. And by extension, doesn't that show I was just wrong to think in S1 that it was permissible for me to flip the switch?

2. We would be more inclined to accept that my flipping the switch in S1 was impermissible if we could see a way of fitting that claim into a more general moral theory we found plausible. So let us see what kind of theory that might be.

The following seems very plausible.

Death Thesis (DT) Other things being equal, X ought not to cause Y's death.

(Here and elsewhere 'X' and 'Y' range over persons.)

What is quite crucial here is the 'other things being equal' clause. Now other things are clearly not equal when Y is about to murder X or X's children, and can only be stopped by X's killing him. More controversially, some people think that other things are at least sometimes not equal when X can bring about some great amount of good, such as saving the lives of five people, by causing Y's death. But these and similar circumstances have no bearing on the S variations.

But are other things also not equal when there is some action which will cause Y's death which X does not believe and has no reason to believe will cause Y's death? One reason for not allowing the 'other things being equal' clause to extend that far is that it would then be harder to derive the claim that it was impermissible for me to flip the switch. Since that claim is supported by the argument from the objectivity of ought, that argument supports not allowing the

'other things being equal' clause to extend that far. But there is also a second reason. Suppose there is a button wired to an incendiary device on Smith's back with: "Press me to incinerate Smith," written on it. If anything is impermissible, my pressing that button is. And the simplest explanation, and therefore, other things being equal, the best explanation, is that I ought not to cause Smith's death, despite the fact that there is some other action (my flipping the switch) which will cause Smith's death which I neither believe nor have reason to believe will cause Smith's death.

Now I do not want to say that it would be flat out inconsistent to say that other things were not equal in S1. I only want to establish the weaker claim that it is reasonable to deny that other things were not equal in S1. Hence it is reasonable to accept that it follows from DT that it was true at the time of my flipping the switch that

(1) I ought not to cause Smith's death.

How do we derive from this that it was impermissible for me to flip the switch? The following inheritance principle[5] looks very plausible.

IP1 If X's doing F will cause Y's death, and if X ought not to cause Y's death, then X ought not to do F.

My having caused Smith's death by flipping the switch showed that it was true at the time of my flipping the switch that

(2) My flipping the switch will cause Smith's death.

It follows from (1), IP1, and (2) that it was true at the time of my flipping the switch that I ought not to flip it; or equivalently: my act of flipping the switch was impermissible. Since it is at least reasonable to accept DT, the claim that DT entails (1), IP1, and (2), this gives us what we wanted, a plausible moral theory into which we can fit the claim that it was impermissible for me to flip the switch.

3. There is even another moral phenomenon which seems to support the separation of acts from agents and the conclusion that I acted impermissibly in flipping the switch.

When I discover that I have caused Smith's death, many people think it morally appropriate that I feel regret. Suppose there was someone standing next to me as I flipped the switch, and he dis-

covered that I had thereby caused Smith's death at the same time as I discovered it. It is commonly thought that although it is morally appropriate for him to feel some regret, it is morally appropriate for me to feel much more intense regret.

What explains the asymmetry? It cannot be that I, unlike the bystander, was at fault in S1: I was not. One suggestion is that it is morally appropriate that my regret be much more intense because I, unlike the bystander, acted impermissibly in causing Smith's death.[6]

4. Let us review what we have. In the introduction we saw a general motivation for the separation of acts from agents. S1 is just the kind of case on which that doctrine bears. The argument from the objectivity of ought strongly supports the claim that I acted impermissibly in S1, and we have seen a plausible looking moral theory – DT together with IP1 – from which it is reasonable to think that claim can be derived. Furthermore, the claim appears to explain the appropriateness of agent regret. When all this is taken together, the overall picture is of a number of claims which have independent plausibility and which hang well together and provide mutual support. We are uneasy when we think that I was surely justified in flipping the switch, but defenders of the doctrine of separation of acts from agents would tell us that is only because we are overly influenced by the thought that an agent's having acted impermissibly suggests that he was not justified. While that is often true, it is not always true, as can happen in cases like S1 in which reasonable beliefs about what will happen and what does happen come apart dramatically.

5. Many of us should find this picture worrying, however. If risk is not relevant to permissibility, we may be in trouble elsewhere in moral theory.

One class of cases involves indivisible efforts or goods. For example, we can rescue the five people stranded on one island or the one on the other, but not both groups.[7] Or we have just enough drug to save the lives of five people with one disease or one with another, but not both groups. What should we do? Suppose we want to treat them equally. One suggestion is that we roll a six sided die and give the group of five a five in six chance of being saved and the group of one a one in six chance. We could then claim that we were treating

each person equally as each is given an equal chance, with the group of five pooling their chances.[8]

Another case is the Trolley problem.[9] A trolley is hurtling out of control down the track and will take the left hand track, thereby causing the deaths of five people, unless you flip a switch and cause the trolley to take the right hand track, thereby causing the death of a sixth person. Many people believe that at least in some variations of this problem it is permissible for you to flip the switch. The puzzle is then to explain why that is so when in general it is impermissible to kill one person in order to save five.

Judith Thomson suggests that what is distinctive about the Trolley variations in which flipping the switch is permissible is that unlike most other cases in which you can kill one to save five, there was some earlier time at which each of the six was better off for its being permissible for you to kill the one.[10] Details aside, assume that at some earlier time each of the six was equally likely to occupy any position in a trolley situation. Then given that they will find themselves in a trolley situation, its being impermissible to turn the trolley would mean that, at that time, they each face a five in six chance of being killed, while its being permissible to turn the trolley would mean that, at that time, they each face a one in six chance of being killed, a clear improvement for each of them.

I shall not try to argue this, but I strongly suspect that these accounts are threatened by the same considerations that led us to conclude that it was impermissible for me to flip the switch and hence that risk is not relevant to permissibility. If, like me, you find some of these ideas attractive, you will have reason to hope there is a satisfactory alternative to the picture we have just constructed.

6. We can now begin the argument against the claim that it was impermissible for me to flip the switch and, more generally, against the moral picture surrounding the separation of acts from agents. I start with the argument from regret.

Consider Ray Gun:[11] you are at the bottom of a well and a villain throws a young man down. He will land on you, and you will be killed, and he will live, unless you fire your ray gun at him, disintegrating him, and thereby killing him and saving your life.

If you kill him it seems morally appropriate for you to feel intense regret, more regret than it would be appropriate for an uninvolved bystander to feel. But we cannot say that what explains this is that it was impermissible for you to fire the ray gun: it was clearly permissible. Hence we cannot infer from the appropriateness of my feeling intense regret in S1 to the claim that I acted impermissibly.

Furthermore, we can explain why (we think) regret is appropriate in both cases. People who are disposed to feel intense regret whenever they cause death, permissibly or otherwise, may be more strongly disposed to act permissibly, and in seeming to place more value on life they may be in many ways better to live with.

Some may think regret not called for in Ray Gun. They, like Lady Macbeth, may think it a waste of time: 'Things without all remedy should be without regard: what's done is done,'[12] and that morality does not call on us to waste our time. But then they have given us no reason to think that regret is called for in S1: they cannot say regret is called for because, unlike in Ray Gun, I acted impermissibly, because they want to use the claim that regret is called for in S1 to infer that I acted impermissibly.

7. Rejecting the argument from regret does not, however, show that it was permissible for me to flip the switch in S1. It merely undercuts a reason for denying it.

It is natural to argue that it was permissible for me to flip the switch in S1 by arguing that risk is relevant to the permissibility of actions. And it is natural to do that by considering an example like the following.[13] Jones is unconscious and about to lose his left hand unless you operate, imposing on him a one in a hundred risk of death. You happen to know that Jones has recently paid to have an operation to save his left hand for which he knowingly bore a one in a hundred risk of death. So there is plenty of evidence that Jones in sound mind prefers bearing a one in a hundred risk of death to losing his left hand for certain. Your operating seems intuitively permissible, but suppose he dies. Then it is open to a defender of the separation of acts from agents to say that his later dying entails that it was impermissible for you to operate, despite your having been perfectly justified in operating.

To get around this problem consider

R(oom) 1 There are a hundred temporarily unconscious people in a room. Each will have his left hand chopped off unless you pull the trigger of the gun that is pointed at one of them, Jones, thereby causing his death.

It is intuitively clear that it is impermissible for you to pull the trigger. Indeed this claim follows from the fragment of moral theory we have built around DT. The good that would result from your pulling the trigger, although significant, is not sufficiently great for this case to be covered by the 'other things being equal' clause of DT, and so on.

But now consider

R2 The same as R1, except you have a third option: press the red button and a gun will spin at random and fire when it stops spinning, thereby causing the death of the person it is pointed at but saving the left hands of the others. Each person will face a one in a hundred chance of being the person killed.

Pressing the red button imposes a one in a hundred risk of death on each person and gives him a ninety-nine in a hundred chance of having his left hand saved. As before, let us assume that there is plenty of evidence that each in sound mind prefers bearing a one in a hundred chance of death to losing his left hand.

When you press the button the gun ends up being fired at Jones, thereby causing his death. So at the time of your pressing the button the following was true. Your pressing the button will cause Jones' death.

Poor Jones! But does the fact that the outcome of your pressing the button in R2 is exactly the same as the outcome that would have resulted from the impermissible act of your pulling the trigger in R1 entail that you acted impermissibly in pressing the button?

Surely not. To see just how implausible that conclusion would be, consider the fact that there was nothing special about Jones. No matter who had ended up being killed, a similar argument would tell you that your pressing the button was impermissible. So you were not even justified in pressing the button, you were at fault in pressing it, and so on. But that is wildly implausible.

What explains why it was permissible for you to press the button is that pressing it was to the *expected* benefit of each person in the room, including Jones: each person was prospectively better off bearing a one in a hundred chance of death and having a ninety nine in a hundred chance of not losing his left hand than losing his left

hand for certain. By contrast, the expected benefit to Jones in R1 was not sufficiently great to make it permissible to pull the trigger.

One might say that what made it permissible for you to press the button in R2 is that everyone would have consented to your pressing it. But what explains that is just the fact that it was to each person's expected benefit. In my view, an appeal to one form or another of hypothetical consent is not doing any real work here, since what is doing the work is that it was to each person's expected benefit. But in any case, whether you go to the permissibility of pressing the button directly from the facts about expected benefit, or whether you pass through hypothetical consent along the way, R2 shows that the risk an action imposes really is relevant to permissibility. And if that is so, it surely follows that in flipping the switch in S1 I acted permissibly: the risk was trivial.

8. What does accepting that risk is relevant to permissibility mean for the separation of acts from agents? Recall that in S1 the risk of death my flipping the switch would impose was low, but that in S2 the risk was high. But the wiring was the same in each case. The only difference was in the evidence I had. Hence we should say that the risk an agent imposes in performing a particular action is relative to the evidence the agent has.

But now consider

S4 Same as S1, except that it is Bloggs the electrician who is about to flip the switch. She sees the same things I saw in S1, but since she is an electrician she can tell the wiring is in a dangerous state.

I think we should say that Bloggs' flipping the switch in S4 would impose a high risk of death on Smith. But the only relevant difference between S1 and S4 is the difference in the information Bloggs and I have: she is an electrician, I am not. Hence we should say that the risk an agent imposes in performing a particular action is relative to the information the agent has, where information is understood as something like reasonable belief.

Combining this: the risk an agent imposes in performing a particular action is relative to the evidence and information that is available to the agent. Or as we might say: the relevant probabilities are agent-relative.

Does it follow that the probabilities are purely subjective? In other words, that the risk an agent imposes is just whatever he thinks he imposes? No. If in S2 I had believed my horoscope and concluded that my flipping the switch would not impose a high risk I would have been mistaken. Similarly, if in S4 Bloggs had because of her horoscope believed she would not be imposing a high risk, she would have been mistaken. Thus although the relevant probabilities are agent-relative, they are not subjective: they vary objectively with the evidence and information that is available to the agent.

Since the permissibility of at least some of an agent's actions depends in part on the risks they impose, and since the risks they impose are a function of the evidence and information available to him, it follows that the permissibility of some acts depends in part on what it is reasonable for the agent to believe rather than on the way the world really is. That is just what the separation of acts from agents denies, hence we must reject that doctrine.

9. We have seen that the conclusion that it was impermissible for me to flip the switch in S1 follows deductively from DT, our understanding of when other things are not equal, and IP1, in conjunction with the description of the case. We therefore have to revise those premises. We would also like the revised set of premises to provide us with a plausible moral theory which makes risk relevant to permissibility. There are several options which turn out to be more or less equivalent in terms of which risk of death imposing actions are permissible. One might, for example, continue to accept DT and somehow let the 'other things being equal' clause be sensitive to risk. But for technical reasons, and to highlight the role of risk, it is neater to reject DT, and to replace it with

High Risk of Death Thesis (HRDT) Other things being equal, X ought not to impose a high risk of death on Y,

where the relevant probabilities vary objectively with the evidence and information available to X.

One difficulty is that it is far from obvious what 'high' means. But any plausible moral theory faces a similar problem. Whatever your theory, you will have to say that other things being equal, one is not justified in imposing a high risk on others. Since everyone faces

this problem, and since the issues surrounding it would take us some way from our main concern, I suggest we bypass it.

How powerful a principle is HRDT? In other words, how many facts about the morality of impositions of risks of death does it explain? HRDT tells us that we have the right that others not impose high risks of death on us. So this is in part a question about the explanatory power of the best theory of rights, and I don't think we are yet in a position to answer that question, so I am going to have to leave this open.

Another question concerns actions which impose risks of death on others which are not high. It seems to me very plausible that such actions can be morally impermissible. Consider (a) a case where I impose a non-high risk of death on you only because I dislike you and I rather hope this will lead to your death, and (b) a case where for some minor benefit I impose a non-high risk of death on each of ten million people, so that it is very likely that at least one of them will be killed as a result. But it also seems to me very plausible that HRDT cannot explain such facts.

I believe we can explain such facts by accepting

Risk of Death Thesis (RDT) Other things being equal, X ought not to impose a risk of death on Y.

Furthermore, I believe we have no good reason to reject RDT, so we should accept it.[14] But this claim looks vastly implausible, and I cannot begin to argue for it here. So here I will merely explore the consequences of accepting HRDT, with the caveat that it almost certainly does not contain the whole truth about the morality of actions which impose risks of death on others, which is not, of course, to say that it is not true (note that RDT entails HRDT).

How can we derive claims about the permissibility of actions from HRDT? This cousin of IP1 is surely very plausible.

IP2 If X's doing F will impose a high risk of death on Y, and if X ought not to impose a high risk of death on Y, then X ought not to do F.

In S1 I imposed a tiny risk of death on Smith, so rejecting DT and replacing it with HRDT makes it no longer possible to derive that it was impermissible for me to flip the switch. In R2 you imposed what was presumably a high risk of death on Jones and others (a one in a hundred risk on each). But the expected benefit to Jones

and others (a ninety nine in a hundred chance of not losing his left hand for each) of your pressing the button sufficiently outweighed the risk of death. Hence other things were not equal, so it does not follow from HRDT that you ought not to have imposed that risk on Jones and others. Hence it is no longer possible to derive that it was impermissible for you to press the button. By contrast, your pulling the trigger would have imposed a very high risk of death on Jones – a risk with probability one – and the resultant good (ninety nine others not losing their left hands) was not sufficient to outweigh the expected burden of imposing that risk on him. Hence it follows from HRDT together with IP2 and our understanding of the 'other things being equal' clause that in R1 it was impermissible for you to pull the trigger.

More generally, suppose an action which imposes a high risk of death happens to cause death. Applying the 'other things being equal' clause as we have been understanding it in DT means balancing the burden to Y of his death being caused against other factors. Applying the 'other things being equal' clause in HRDT in the same manner means balancing the burden to Y of his bearing a high risk of death against other factors. That is what enables us to derive the right results about R1, R2 and S1. It is easily seen that this also gets us the results we want in the other S variations.

Rejecting DT may look implausible. After all, surely there are a great many occasions on which you ought not to perform an action in virtue of the fact that it will cause someone else's death. For example, you are standing in front of me minding your own business. Surely it is impermissible for me to load my gun, point it at you, and pull the trigger. And isn't that because I ought not to cause your death? But in pulling the trigger I would be imposing an extremely high risk of death on you, so HRDT tells us, quite rightly, that I would be acting impermissibly. More generally, if an agent is reasonably well informed about the likely consequences of his actions, and he intentionally causes someone else's death, then he will also have imposed an extremely high risk of death on that person. DT and HRDT will then almost always agree on whether he acted permissibly.

It is only in cases like my flipping the switch and your pressing the button that there is significant disagreement between DT and HRDT, at least when the other things being equal clause in DT is not

sensitive to risk. But it is about precisely those cases that DT yields the wrong conclusion. Thus in rejecting DT and replacing it with HRDT we are in essence merely discarding some of the unattractive features of DT.

There is even a further advantage to adopting HRDT. Suppose I do pull the trigger, imposing an extremely high risk of death on you. You are lucky: the bullet misses. But intuition suggests that nevertheless I acted impermissibly. That follows straightforwardly from HRDT, but in the absence of that claim it is not obvious how to derive it.

10. We can now consider the strongest seeming objection to the view that what it is permissible for an agent to do depends (in part) on what it is reasonable for him to believe rather than the way the world really is, the argument from the objectivity of ought. To recall, the case which lies at the heart of this argument is:

S3 Same as S1, except as I am about to flip the switch I think to myself: "It is permissible for you to flip the switch." But along comes Bloggs the electrician. She notices that the wiring is in a very dangerous condition, and says: "You ought not to flip the switch."

It seems very plausible to think Bloggs' utterance was true, and doesn't that show that when I thought to myself just prior to her utterance: "It is permissible for you to flip the switch," I was thinking falsely? And doesn't that show by extension that when I was alone and thought to myself in our original case S1: "It is permissible for you to flip the switch," I was thinking falsely? It is this interpersonal use of terms like 'ought' and 'permissible' that makes the case for its having been impermissible for me to flip the switch in S1, and more generally, the case for the separation of acts from agents, strongest.

But consider a variant of S3:

S3′ Same as S1, except as I am about to flip the switch I think to myself: "It is permissible for you to flip the switch." But along comes Bloggs the electrician. She notices that the wiring is in a very dangerous condition, and says: "The wiring is in a dangerous condition. You ought not to flip the switch."

Again, just as in S3, it seems very plausible that Bloggs was speaking truly when she said: "You ought not to flip the switch."

Let t be the time at which I complete my thought, and let t′ be the slightly later time at which Bloggs completes her utterance: "The

wiring is in a dangerous condition." Now relative to the evidence available to me at t, my flipping the switch would impose a very low risk of harm on Smith. Hence it does not follow from our theory that at t it was impermissible for me to flip the switch. But when Bloggs finishes saying: "The wiring is in a dangerous condition," I have a new piece of evidence, namely that Bloggs, who is an electrician, is apparently sincerely asserting that the wiring is in a dangerous condition. Hence at t′ relative to the evidence and information that is available to *me* the probability of Smith's being killed is high. So when, moments after t′, Bloggs says: "You ought not to flip the switch," she is speaking truly. But that is consistent with the proposition I expressed by thinking to myself at t: "It is permissible for you to flip the switch." In short: it is (tenselessly) permissible for me to flip-the-switch-at-t, but not permissible for me to flip-the-switch-at-t′.

And don't we have to say this? Imagine Cook were the coroner determining the cause of Smith's death in S1. Cook discovers that the wiring was in a dangerous condition, but she would be mistaken to think that it was impermissible for me to flip the switch when I did. Similarly, in S3′ Bloggs would have been mistaken in thinking that it was impermissible for me to flip the switch at t. But it does not follow that after she has presented me with new evidence that it is permissible for me to flip the switch at t′.

More generally: since we can learn about the world over time without the relevant parts of the world changing, and since talk of risk is sometimes a reflection of our ignorance, if the permissibility of our actions depends in part on the risks they impose, what is in some sense the same action may be permissible at one time but become impermissible at a later time.

We can now turn to the most compelling case for the view that it was impermissible for me to flip the switch in S1, namely S3, where Bloggs says, without preamble: "You ought not to flip the switch." But I suggest that Bloggs' assertion is true, and that this is consistent with its having been permissible for me to flip the switch before her assertion. Why? I suggest that we understand Bloggs to be asserting something like the following: "Some state of affairs obtains such that if you were aware of it, it would be true that you ought not to flip the switch." Again, that gives me a new piece of evidence which I am

required to take into account, and given that Bloggs is an electrician and seems sincere, after her assertion it becomes true that I ought not to flip the switch, at least not without asking her why she said what she said. Roughly speaking, what is communicated by Bloggs' assertion makes its literal construal true. (And if she hadn't made that assertion, its literal construal would not have been true.)

Contrast S3″, which is like S3 except that it is not Bloggs but Bloggs' child, who knows nothing about anything, who says: "You ought not to flip the switch." That doesn't give me any reason to believe the wiring is in a dangerous condition, or anything else for that matter. In S3″ it is still permissible for me to flip the switch after the child's assertion.

11. We have been looking at causings of death and risks of death, but it is plain that all the arguments would have gone through almost unchanged if 'death' were replaced by 'harm.' Thus I shall have taken us to have argued that we should reject

Harm Thesis (HT) Other things being equal, X ought not to cause Y harm,

and replace it with

High Risk of Harm Thesis (HRHT) Other things being equal, X ought not to impose a high risk of harm on Y.

In fact, I believe we should accept the even stronger

Risk of Harm Thesis (RHT) Other things being equal, X ought not to impose a risk of harm on Y.

But I can't begin to argue for this here,[15] so I'll stick to the weaker HRHT.

Now HT and HRHT just are, respectively, the theses that Y has a (claim) right against X that X not harm her and that Y has a (claim) right against X that X not impose a high risk of harm on her. Since DT, and more generally, HT, are generally regarded as being part of our best theory of rights, I am suggesting a revision in that theory. It is worth noting some of what this amounts to.

I have argued that the risk an agent imposes in performing a particular action is a function of what it is reasonable for him to believe rather than the way the world really is. Hence, according to HRHT, the permissibility of some of his actions is a function

in part of what it is reasonable for him to believe rather than the way the world really is. This provides a degree of support for the claim that if X acts impermissibly then X is at fault. (I only say 'a degree of support' because that claim is very much more general than I know how to argue for.) One might think that it provides a degree of support for the claim that if X infringes someone's right then X is at fault. It does not. Sometimes rights infringements are straightforwardly permissible:[16] consider the case in which I punch you on the arm to save five lives. That is one reason why HT and HRHT have 'other things being equal' clauses.

12. There is a lot more that could be said about risk, but we are exploring the relevance of an agent's beliefs to the permissibility of her actions, and I want to mention a different kind of case.

It is common to distinguish between risk and uncertainty. When an agent knows the probabilities of the various possible outcomes she is said to be acting under conditions of risk. When she does not know the probabilities of at least some of those outcomes, she is said to be acting under conditions of uncertainty.

This is a rough distinction, and it depends on a particular view of probabilities. On some subjective accounts there is no such distinction. But the kinds of probabilities that are relevant to the moral problems we have been discussing are objective, and on any plausible account of what it means to be an objective probability, there will be circumstances in which there is too little relevant evidence or information for objective probabilities to be well defined.[17] In such circumstances moral problems will arise which are analogous to the problems of decision making under conditions of uncertainty.

It is far from obvious what to say about the permissibility of actions whose consequences are uncertain. At least in the context of actions which impose risks on others, well developed machinery exists which provides, for example, a way of describing the magnitudes of the risks. I am far from certain that I have anything useful to say about problems involving uncertainty other than to draw attention to their existence. But one general remark seems reasonable. An important, if rough, distinction between circumstances in which objective probabilities are well defined and those in which they are not is that obtaining new evidence will often be more informative in

the latter case than in the former. (Compare the possibility that taking aspirin will produce some undesirable side-effects with the possibility that taking some newly discovered and untested drug will.) Thus although it is relevant in the case of risk imposition, the possibility that one should obtain further evidence before performing an action whose consequences are potentially harmful is especially important when those consequences are uncertain. The more harmful some of the possible consequences, the less the cost to the agent of obtaining new evidence, and the more likely that evidence is to be informative, the greater the reason for thinking that the action in question is impermissible unless the agent first obtains that evidence.

13. In conclusion I shall make three more general remarks on making the permissibility of an agent's acts depend (in part) on the beliefs it is reasonable for her to have rather than on the way the world really is.

First, it may make it easier for moral theory to be action guiding in an attractive sense, and it may provide a more attractive account of the postulate that ought implies can. When acts are separated from agents, moral theory is not action guiding in the sense that a fairly well informed and normally reasoning agent can always determine what she is permitted to do. But by making the permissibility of her actions depend in part on the beliefs it is reasonable for her to have rather than facts about the way the world really is, such an agent can always determine what she is permitted to do, and we can give an epistemic sense to the 'can' in the claim that ought implies can.

Second, it may make us more willing to make the permissibility of an agent's acts also depend in part on other facts about her mental life. It is easy to feel that if, as the argument from the objectivity of ought claims, an agent's beliefs are not relevant to the permissibility of her actions, then neither are other facts about her mental life. Answering that argument may add to the plausibility of, for example, letting an agent's intentions be relevant to the permissibility of her actions. It may also motivate making the permissibility of her actions depend in part on how well she is able to reason. Reasonable beliefs are often characterized in terms of the reasoning capacities of normal agents. But that may exhibit an unreasonable bias against the less intelligent. In making permissibility depend on the evidence available to an

agent we would not, after all, insist that visual evidence is evidence available to people who are blind.

Third, it may pave the way to a more unified moral theory; in particular, a more unified account of rights and virtues. It is not at all plausible to suppose that all facts about the permissibility and impermissibility of actions can be explained by a theory of rights. Many writers have thought that some such facts are best explained in terms of their relations to various virtues. Thus Thomson suggests, surely correctly, that we should take seriously

> *Thesis of Moral Requirement* A person is morally required to do a thing just in case his or her refraining from doing the thing would be morally bad in some way–mean or cowardly or unjust and so on.[18]

And, she says, if a person is morally required to do a thing, then he or she ought to do the thing. Hence a person ought to do a thing if refraining from doing the thing would be mean or cowardly or unjust.

Here an interesting tension arises. Whether an agent's act is generous or brave (and, consequently, mean or cowardly) depends on what the agent reasonably believes. It is, for example, brave only if the agent reasonably believes she is putting herself at risk for the sake of an outcome that is good in some way, generous only if she reasonably believes she is incurring a cost for the sake of others.[19] So the fragment of morality connected with generosity and bravery in action is best understood as denying that what an agent ought to do is independent of her reasonable beliefs.

What is it for an act to be unjust? It is for it to be a rights violation.[20] There would be an odd tension in the Thesis of Moral Requirement if when it came to the fragment of morality connected with rights infringements, what an agent ought to do *is* independent of what it is reasonable for her to believe. What on earth would make for this difference between, on the one hand, generosity and bravery in action, and on the other, justice in action? Thus in denying the separation of acts from agents in the context of rights infringements, we are making the Thesis of Moral Requirement smoother, and this may pave the way to a more unified understanding of the relation between justice in action with things like generosity and bravery in action, but I'm going to have to leave this as a conjecture.

NOTES

* I am very grateful to Frank Arntzenius, Barbara Herman, and Kadri Vihvelin for generous and extremely useful comments on earlier versions of this paper. I have also benefited from helpful comments by a referee for *Philosophical Studies*.

[1] Moore (1958, 118–121).

[2] A similar example is discussed in Thomson (1986a). In saying how much there is to be said for what I am calling the doctrine of separation of acts from agents, I am deeply indebted to that article, as well as to a later discussion of similar issues in Thomson (1990, 187–191, 227–234, 239–242).

[3] Moore (1958, 121).

[4] Thomson describes this argument, and cautiously suggests that it is very hard to see what is wrong with it, in Thomson (1986a).

[5] The term 'inheritance principle' is from Thomson (1986a) where a number of such principles are discussed.

[6] For discussion of inferences about the moral status of actions from the later appropriateness of the agent's regret see Williams (1973); Williams (1981); and Thomson (1990, 96–97, 240–242).

[7] The importance of these kinds of cases was brought to life in Taurek (1977).

[8] See Kamm (1991) where this idea of using proportional chances is defended in at least some kinds of cases, and the connection with treating people equally is described.

[9] From Foot (1978).

[10] See Thomson (1990, chapter 7).

[11] Based on Nozick (1974, 34–35).

[12] Shakespeare, *Macbeth*, Act 3, Scene 2.

[13] This example is based on Thomson (1991, 187–191).

[14] See McCarthy (1997). Note that accepting RDT avoids the problem of saying how high 'high' is.

[15] See ibid. for an argument for RHT.

[16] I take this to have been convincingly argued for in Thomson (1986b) and Thomson (1990).

[17] Compare the rejection of the use of the principle of sufficient reason in certain moral contexts in Rawls (1971, 168–173) in light of the conception of objectivity given in Rawls (1993, 110–112).

[18] Thomson (1996, 151).

[19] Ibid., 145. I say 'reasonably believes' rather than just 'believes' because, for example, if I sincerely but insanely believe that by writing this sentence I am saving the world from a great evil but incurring a huge risk of assassination, it is wrong to say that I act bravely in writing this sentence.

[20] Ibid., 146–147, esp. n. 13.

REFERENCES

Foot, P. (1978): 'The Problem of Abortion and the Doctrine of Double Effect', in *Virtues and Vices*, Los Angeles: University of California Press.

Kamm, F. (1991): *Morality, Mortality*, Volume 1, New York: Oxford University Press.
McCarthy, D. (1997): 'Rights, Explanation, and Risks', *Ethics* 107, no. 2.
Moore, G. (1958): *Ethics*, London: Oxford University Press.
Nozick, R. (1974): *Anarchy, State, and Utopia*, New York: Basic Books.
Rawls, J. (1971): *A Theory of Justice*, Cambridge: Harvard University Press.
Rawls, J. (1993): *Political Liberalism*, New York: Columbia University Press.
Taurek, J. (1977): 'Should the Numbers Count?', *Philosophy and Public Affairs* 6, 293–316.
Thomson, J. (1986a): 'Imposing Risks', in Thomson (1986b).
Thomson, J. (1986b): *Rights, Restitution, and Risk*, Cambridge: Harvard University Press.
Thomson, J. (1990): *The Realm of Rights*, Cambridge: Harvard University Press.
Thomson, J. (1996): 'Moral Objectivity', in Harman, G. and Thomson, J. (eds.), *Moral Relativism and Moral Objectivity*, Oxford: Blackwell.
Williams, B. (1973): 'Ethical Consistency' in *Problems of the Self*, Cambridge: Cambridge University Press.
Williams, B. (1981): 'Moral Luck', in *Moral Luck*, Cambrodge: Cambridge University Press.

Program in Law, Ethics, and Health
School of Hygiene and Public Health
Johns Hopkins University
Baltimore, MD 21205-1996
USA