A Consequentialist Distinction between What We Ought to Do and Ought to Try

INGMAR PERSSON
Gothenburg University

G. E. Moore raised the question of whether consequentialists ought to maximize actual rather than expected value, and came down in favour of the former alternative. But rather recently Frank Jackson has presented an example which has been widely thought to clinch the case in favour of the alternative view. This article argues for a sort of compromise between these rival views, namely that while we ought to *do* what maximizes actual value, we ought to *try* to do what maximizes expected value. It is claimed that consequentialists could consistently adopt this view, though in Jackson's case they are certain that, if they try to maximize expected value, they shall most likely not maximize actual value.

In *Ethics* G. E. Moore raised the since much-debated question of whether consequentialists should claim that we ought to maximize actual or expected value.[1] Moore came down in favour of the former alternative, but rather recently Frank Jackson has presented a sort of example which has been widely thought to clinch the case in favour of the latter alternative:

> Jill is a physician who has to decide on the correct treatment for her patient, John, who has a minor but not trivial skin complaint. She has three drugs to choose from: drug A, drug B and drug C. Careful consideration of the literature has led her to the following opinions. Drug A is very likely to relieve the condition but will not completely cure it. One of the drugs B and C will completely cure the skin condition; the other though will kill the patient, and there is no way that she can tell which of the two is the perfect cure and which the killer drug. What should Jill do?[2]

Imagine that the actual value of a complete cure is 50, that of a partial cure 40 and that of death −100. Let us also assume, for the sake of simplicity, that the probability of A producing a partial cure is 1, and the probability for each of B and C that they produce a complete cure is 0.5 and the probability that they produce death is the same. The expected value of Jill's prescribing A is then 40 and −25 both of her

[1] G. E. Moore, *Ethics* (Oxford, 1966 [first published 1912]), pp. 98–101.

[2] 'Decision-theoretic Consequentialism and the Nearest and Dearest Objection', *Ethics* 101 (1991), pp. 462–3. In an unpublished paper by Derek Parfit, 'What We Together Do' (1988), there is a similar sort of case. Parfit refers to Donald Regan as inspiration.

© 2008 Cambridge University Press
doi:10.1017/S0953820808003191
Utilitas Vol. 20, No. 3, September 2008
Printed in the United Kingdom

https://doi.org/10.1017/S0953820808003191 Published online by Cambridge University Press

prescribing B and of her prescribing C. For the sake of completeness, let us add that, if Jill does not act now, John's disease is certain to be incurable, with a value of 0. Jill will then be maximizing expected value if she prescribes A, but she knows for sure that what will maximize actual value is prescribing either B or C.

It seems very plausible to claim that what Jill should or ought to do in this situation is to prescribe A. But then it follows that what she ought to do is to maximize expected rather than actual value. Now, my answer will be something of a compromise between these two rival views. For I shall argue that Jill ought to *try* to prescribe A, but that she ought to prescribe the drug that will cure John completely. We may assume that, unbeknownst to her, this is drug B. Generally speaking, in some situations we ought to try to maximize expected value, though we are certain that this will not maximize actual value, which is what we ought to *do*. These are situations in which it would be irresponsible and blameworthy to try to maximize actual value because there is an unacceptable risk that we would thereby cause serious harm.

In support of this distinction, suppose that Jill decides or chooses to prescribe A, writes down the prescription on her computer and sends off an e-mail with this prescription. During the transmission of her message, however, there occurs some kind of totally unforeseeable failure so that the message received is a prescription of B. Consequently, John gets B and completely recovers. It seems clear that in this case, *Lucky Failure*, Jill has not failed with respect to anything she ought. She decided and tried to prescribed A, and this is the responsible course to take. It would have been irresponsible of her to choose between B and C, since in order to have a 50 per cent chance of slightly improving John's situation, she would have run an equally great risk of killing him. So, she decided, and tried, to do what she ought to decide and try to do. But, owing to totally unforeseeable fortuitous circumstances, she has in fact achieved what she wished she could achieve, to prescribe – or at least bring it about that he gets – the drug which causes John's complete recovery.

Surely, there is no reason to regret that this transmission failure occurred, though there would have been such a reason if the transmission failure had caused her to fail to do something that she ought. Instead, it seems that Jill, and we, ought to have hoped or wished for the occurrence of some such unexpected event which would make her do what has in fact the best consequences. If so, it cannot be that Jill ought to have prescribed A to John (if we take this to entail bringing it about that John gets A). As Jackson himself concedes,[3] after we

[3] Jackson, 'Decision-theoretic Consequentialism', p. 471.

https://doi.org/10.1017/S0953820808003191 Published online by Cambridge University Press

have discovered that B is the drug which completely cures John, it is plausible to say that this is the drug Jill ought to have prescribed.

Thus, *Lucky Failure* supports the claim that Jill ought to (decide and) try to prescribe (and bring it about that John gets) A, which is what maximizes expected value, though she ought to prescribe (and bring it about that John gets) B, which is what maximizes actual value. Hence, we sometimes ought to *try* to maximize expected value, though we are convinced that this will not make us maximize actual value, which is what we ought to *do* and, so, we cannot in these situations regard ourselves as trying to do what we ought to do. Because of the happy transmission failure, Jill succeeds in doing both what she ought to try to do and what she ought to do. So, she will have reason to hope or wish that some unexpectable event of this sort occurs. But she surmises that it most certainly will not, and that she will therefore fail to do what she ought to do in this situation, by trying to do what she ought to try to do.

These situations in which we believe that trying to do what we ought to try to do is incompatible with doing what we ought to do, and in which we thus ought to hope or wish that we will not succeed in doing what we try to do, are exceptions. The rule is rather that the option that will maximize expected value will be the same as the option that we think will maximize actual value. True, we are usually aware that the evidence we could assemble is incomplete. But we can rarely tell that, according to the complete, unavailable evidence, another act than the one which has come out as maximizing expected value on our evidence will come out as the one that will maximize actual value. In the normal circumstances, then, trying to do what maximizes expected value and trying to do what maximizes actual value will coincide, and we have no reason to hope that we shall fail to do what we try to do.

It might however be thought that it *cannot ever* be true that we ought to hope that we shall fail in trying to do what we ought to try to do, and thereby do what we ought to do. For it might be thought that when we ought to do something, we ought to do it *intentionally*,[4] where this implies trying and succeeding. Certainly, it could not be true both that Jill ought to try to prescribe A intentionally and that she ought to prescribe B intentionally, and thus successfully attempting to do so, given that, as we are assuming, she cannot try to prescribe both A and B.

But it has awkward consequences to hold that whenever we ought to do something, we ought to do it intentionally. Imagine that I ought to kill the Terrrorist. I take my gun and drive off to his lodgings, intending to shoot him dead. On my way through town, I accidentally run down

[4] See Michael J. Zimmerman, 'The Relevance of Risk to Wrongdoing', *The Good, the Right, Life and Death*, ed. Kris McDaniel, Jason R. Raibley, Richard Feldman and Michael J. Zimmerman (Aldershot, 2006), p. 162.

https://doi.org/10.1017/S0953820808003191 Published online by Cambridge University Press

and kill a pedestrian who happens to be the Terrorist. If it had been true in this case, *Terrorist Killing*, that I ought to kill him intentionally, I would not have done what I ought. I would instead have reason to fear such a mishap, since it would make it impossible for me to do what I ought. This seems absurd, however, for what I ought to do is simply to kill the Terrorist, and not to kill him in any particular way, such as by shooting him. But if we suppose that I ought to kill him intentionally, whether I succeed in doing what I ought will come to depend upon whether I actually apply my intended means of killing. For I cannot intend and try to kill somebody without intending or trying to kill him in some particular way. So, whether I succeed in doing what I ought will come to depend upon whether I succeed in acting in some particular way in which I am not obliged to act, for otherwise my act will not be intentional. In contrast, on my view I have done everything I ought in *Terrorist Killing*, for I tried to kill the Terrorist, and I did in fact. It does not matter that the way in which I tried to kill him differs from the way in which I in fact killed him and, so, that I did not kill him intentionally.

When at least some consequentialists claim that we ought to maximize actual value, they do not require that we ought to do this intentionally, or try to do it. For they stress that they are not putting forward a 'decision procedure', but simply a 'criterion' for when acts are right or ought to be done. Such a criterion does not imply anything about the praise- and blameworthiness of agents, as Moore's emphasizes.[5] It does not imply that, if we fail to do what maximizes actual value, we can justifiably be blamed for it, e.g. that Jill could be blamed for failing to prescribe the drug which completely cures John, though in some sense she could have done so. Conversely, if Jill were to choose between B and C, and happened to select B, she would in fact do what she ought to do, but she can still be justifiably blamed, since she irresponsibly took an unacceptable risk of causing his death.

We could roughly distinguish between the 'internal' and the 'external' side of an action which is intentional under some description. The internal side is what we decide, intend and try to do, while the external side is what we in fact do, intentionally or unintentionally, what changes we in fact cause in our attempt. When we speak of what we ought to do in the context of deliberation, it is the internal side of action we regulate, whereas a criterion of rightness regulates the external side. Our action is intentional when there is a certain (philosophically controversial) 'fit' between the internal and the external side. But, as *Terrorist Killing* is meant to illustrate, if we do what we ought, internally speaking – i.e. try to kill the Terrorist – and we in fact do

[5] Moore, *Ethics*, pp. 100–1.

https://doi.org/10.1017/S0953820808003191 Published online by Cambridge University Press

what we ought, externally speaking – i.e. actually kill the Terrorist – it does not matter from a moral point of view whether the fit which makes the action intentional happens to be absent.

It might, however, seem peculiar to claim that Jill ought to *try* to prescribe A rather than that she ought to prescribe it. To say that someone ought to try to do something, or tries to do it, normally implies uncertainty about success. Thus, I would say that I ought to try to do X instead of that I ought to do X when I doubt whether I have the ability or opportunity to do X. But this uncertainty implication of 'try' is a Gricean implication which could be cancelled by the addition of a suitable disclaimer. For there is nothing contradictory about saying e.g. 'That's what I'm trying to do, and I'm sure I shall succeed'. I stipulate that in the present context 'try' and 'ought to try' should imply no uncertainty about success. For Jill could be taken to have excellent reason to believe that she will succeed in what she is trying.

But suppose that I predict that if I try to do X, I shall most certainly do X. It might then be wondered whether I could consistently hold that I ought to try to do X if I deny that I ought to do X. If this is right, Jill could not consistently hold that she ought merely to try to prescribe A, and not actually to prescribe it. We can however see that there is no inconsistency here by looking at cases in which we believe that we shall most certainly *fail*. For instance, I might consistently think that I ought to try to save my child who has been lost in a vast wilderness, though I am utterly convinced that, when I try, I shall most certainly fail and look only in places in which she is not to be found. It does not follow that I am committed to thinking that I ought to look only in places in which she is not to be found. Therefore, the facts that I believe that I ought to try to do X and that I am certain that my attempt to do X will succeed do not logically commit me to believing that I ought to do X. It is just that one common reason for having this belief – the reason of uncertainty – is absent.

In the midst of the process of deliberating and trying, we cannot contrast what we are trying to do with what we are (unintentionally) doing (just as when we are in the state of believing we cannot contrast what we (falsely) believe with what is the case). So, when I speak of what we ought to try to do, this covers what we see ourselves as doing in these first-person, present-tense cases. Other people can of course contrast what we are trying to do with what we in fact are doing, and so can we *post factum*: we can say e.g. that we tried to prescribe A, but failed and (unintentionally) prescribed B instead. In my talk about what we ought to do, 'do' has this sense which does not imply trying or intending to do what we do. This sense cannot figure in deliberation; therefore, when I speak of what we ought to do, this is not meant to give deliberative guidance or advice.

https://doi.org/10.1017/S0953820808003191 Published online by Cambridge University Press

Now, when we decide what to (try to) do, we ought to do so on the basis of the reasons that are available to us to the extent that we can determine what they count in favour of. We ought to decide and try to do what on this basis comes out as having the highest value. This is so, even if we are convinced that there *are* reasons, or relevant facts, inaccessible to us, according to which some other alternative has the highest value. It would often be irresponsible and blameworthy to rely upon guesses about the purport of these reasons. Jill's situation is of this sort, and she makes the decision that she ought to make when she decides to (try to) prescribe A. The fact that she is convinced that there are reasons which make it true that she ought to prescribe B or ought to prescribe C does not affect what decision she ought to make, since she does not know which alternative they count in favour of.

This distinction between two kinds of reasons explains why one's judgement that one ought to try do X and one's conviction that one most certainly will succeed do not commit one to denying that one ought to do something else, Y. The latter 'ought' is determined by the reasons there are rather than the reasons whose force one is in a position to assess which determine the former 'ought' whose object is trying to do instead of simply doing.

It might be asked why, instead of distinguishing between what we ought to try do and what we ought to do, we could not simply do as some have done and distinguish between what we ought in an 'evidential' sense, which is to maximize expected value, and what we ought in a 'factual' sense, which is to maximize actual value.[6] It is however unsatisfactory merely to distinguish between different senses of 'ought'. For if we were to hold that in one sense, which is relative to the reasons or evidence available to her, Jill ought to prescribe A, but in another sense, which is relative to all facts or everything there is to know, she ought to prescribe B, we shall wonder whether she *really* ought to prescribe A or B, which sense of 'ought' has priority in this conflict. If this question has no answer, Jill will see herself as facing a genuine moral dilemma in which she cannot do everything she ought, cannot avoid acting wrongly in one sense, for we are assuming that she cannot prescribe both A and B.[7] Similarly, in *Lucky Failure* it is unclear whether Jill would have reason to wish for the transmission failure, since irrespective of whether it occurs, she will fail to do something that she ought.

I avoid this dilemma by adding that what Jill ought in the evidential sense is to try to prescribe A, while what she ought in the factual sense

[6] See e.g. Derek Parfit, 'On What Matters' (unpublished MS, 2008), sect. 15. In fact, Parfit recognizes further senses alongside the two mentioned.

[7] Cf. Michael J. Zimmerman, 'Is Moral Obligation Objective or Subjective?', *Utilitas* 18 (2006), p. 332.

https://doi.org/10.1017/S0953820808003191 Published online by Cambridge University Press

is to prescribe B. It is now clear to Jill what she ought to try to do, namely to prescribe A. But, as *Lucky Failure* illustrates, with a bit of luck she can fulfil both of her 'oughts'; so, she has reason to wish for the transmission failure which makes her achieve this. To spell out this proposal in somewhat greater detail, let us assume, as I believe to be true, that 'I ought to . . .' means something like 'I have most (or strongest) reason to . . .'.[8] If so, Jill could claim something along the following lines: 'To the extent that I can tell what the relevant reasons favour, I ought or have most reason to try to prescribe A, but according to all the relevant reasons there are, I ought or have most reason either to prescribe B or to prescribe C, so I hope that I shall unintentionally prescribe whichever of these I have most reason to prescribe.'

We may compare, and contrast, what we ought to do with what we ought to believe. In the latter case, too, we could distinguish between what we ought in the evidential sense – or ought rationally – to believe and what we ought in the factual sense to believe, which is the truth. What we ought to believe in these senses could differ, since the evidence we have may give us most reason to believe something that is in fact false. But in the case of belief there is no distinction between what we try to believe and what we in fact end up believing (though there is a distinction between what we might try to *make* ourselves believe and what we in fact end up making ourselves believe). However, as regards the question of what we ought to believe, a parallel to Jill's situation could not arise: it could not be that, according to the reasons whose force we can appraise, we have most reason to believe p, while we are convinced that the reasons there really are give most support to something incompatible with p. For the latter conviction implies that the reasons it concerns belong to the set whose force we can appraise, in which case these reasons would not after all support p. Hence, we could never coherently see ourselves as facing a conflict between what we ought to believe in these two senses.

I have argued that consequentialists could consistently claim both that we ought to try to maximize expected value and that we ought to maximize actual value, though we occasionally are certain that it is incompatible acts which maximize these values. But this consistency requires that our trying to maximize expected value will make us maximize actual value, not on each and every occasion, but at least more often than something else we could try to do, that it is in general our best means of maximizing actual value. This is, however, a contingent matter: we could imagine that, say, our *luck* had been perfectly reliable, so that whenever we were to make choices

[8] See Appendix.

https://doi.org/10.1017/S0953820808003191 Published online by Cambridge University Press

in ignorance, such as between drugs B and C, we would *always* get what we want. If so, we ought always to try our luck, rather than to try to maximize expected value, since that would then in general be our best means of maximizing actual value. Likewise, we ought to believe what is supported by our best available evidence only because that is in general our most reliable means of believing what is true. If our guesses had been invariably lucky, we ought to have gone by them rather than by the evidence we could collect.[9]

APPENDIX

John Broome views the relation between 'ought' and reasons in a way that is opposite to the one I assume in the text: he defines a (kind of) reason for you to . . . as a fact that explains why you ought to . . . , in a sense of 'explains' which is supposedly not normative.[10] But if the facts that we adduce as our reasons are not normative in the sense that they also *justify* that we ought, it seems that there is not anything that could justify this. There is only a natural explanation of the fact that we ought something. This seems hard to swallow on any metaethical view, but it seems downright incompatible with the view that normative facts are 'non-natural', and we should not accept any analysis of the relation between 'ought' and reasons which begs the question against this view. On the other hand, if Broome were to concede that the facts which explain why we ought to do something also justify it, it seems he has not got rid of reasons after all. Certainly, he could still deny the reducibility of 'ought' that I am assuming, claiming instead that the notions both of 'ought' and of reasons are primary and irreducible. I think this view is a bit extravagant, but I shall not pursue the matter any further. Nor shall I attempt to establish that 'ought' is ambiguous between something like an evidential and factual reading. These are matters I leave behind, as well as the question of what the evidential reading, or the notion of reasons being 'available', precisely amounts to.

Ingmar.Persson@phil.gu.se

[9] For helpful comments I am grateful to participants in a seminar at the Department of Philosophy, Gothenburg University, at which the present article was presented as a paper. But I am particularly grateful to Derek Parfit and Michael Zimmerman.

[10] 'Reasons', *Reason and Value: Essays on the Moral Philosophy of Joseph Raz*, ed. R. Jay Wallace, Philip Pettit, Samuel Scheffler and Michael Smith (Oxford, 2005).

https://doi.org/10.1017/S0953820808003191 Published online by Cambridge University Press