**Not Quite Non-Consequentialism: The Implications of Pettit’s “Three Mistakes about Doing Good (and Bad)” for Metaphysics and Moral Philosophy**

As its title indicates, Philip Pettit’s “Three Mistakes about Doing Good (and Bad)” identifies and rejects three common claims restricting what can count as a good (or bad[[1]](#endnote-1)) effect of action. The key question here is how do we work out how much good you have brought about by your action? The first common claim is that only *causal* effects or consequences of action can count as goods that are brought about by an action. The second, that we can only count *behavioural* effects of action. The third, that we can only count consequences of “an all-or-nothing or on-off or *discrete* kind”. Pettit argues that all three of these claims are mistaken: the first claim is mistaken because our actions can bring about goods that are constitutive consequences of those actions; the second claim is mistaken because the good consequences of an action can depend upon the agent having acted out of certain dispositions; the third claim is mistaken because the good consequences of an action can include the extra effects of pursuing more basic good consequences (such as benefits to others) in a more deliberate or reliable manner. [[2]](#endnote-2)

As Pettit notes, these “mistakes” could be defended in two ways. They could be defended on the basis of a theory of the good, such as Hedonistic Utilarianism, on which there are no goods that can be anything other than all-or-nothing causal consequences of behaviour. Pettit gives examples of constitutive, disposition-based and non-discrete goods.

The second, more interesting defence of the three “mistakes” appeals to a theory of action, which Pettit calls “the production model of action”[[3]](#endnote-3), upon which even if we recognise constitutive, disposition-based and non-discrete goods, these goods cannot be brought about by actions. Pettit offers an alternative model of action, the Control Model, which he argues can make sense of constitutive, disposition-based, non-discrete goods as effects of actions – and is also independently plausible.

Pettit’s paper is designed to have important implications for both metaphysics and moral philosophy. In metaphysics, Pettit uncovers some of the worrying implications of a dominant model of action and provides an interesting alternative. In moral philosophy, Pettit aims to “open up the prospect of reducing the gap between the consequentialist doctrine that acting well is always a function of doing … good and non-consequentialist theories that make it a function of other factors alone or of other factors as well.”[[4]](#endnote-4) The idea is that many of things that non-consequentialists are typically concerned about can count as the good effects of an action. Pettit says: “That makes it much more plausible to think that acting well is a significant function of how much your action thereby improves or promises to improve the world.”[[5]](#endnote-5)

In the next sections, I will give a slightly more fleshed out summary of Pettit’s argument. I will then offer a criticism of Pettit’s argument for the Control Model of Action, suggesting that its initially plausibility is not neutral between different conceptions of the good. Finally, I will try to elucidate Pettit’s claim about the moral implications of his argument and argue that while Pettit is certainly right that his conclusions leave neutral consequentialism much *more* plausible and significantly narrow the gaps between consequentialism and non-consequentialists, we should be careful not to overstate the degree of convergence achieved. Pettit’s neutral consequentialism retains very counter-intuitive implications. Anyone who rejected earlier forms of consequentialism due to *their* counter-intuitive implications is likely to *still* find Pettit’s more sophisticated version implausible.

*The Three Mistakes*

*I. The Only-By-Causal-Consequences Mistake*

The first mistake that Pettit identifies is the view that the only properties of an action that can count as good consequences are causal effects.[[6]](#endnote-6) Pettit gives two types of non-casual properties of an action that can be good-making: conventional properties and circumstantial properties.[[7]](#endnote-7)

Suppose that I return your treasured book at the time I had said I would. By convention, this counts as keeping my promise. The keeping of my promise is not a distinct event caused by my returning the book. There is no temporal lapse between the book being returned and the promise being kept. “The act constitutes or grounds the promise-keeping rather than causing it.”[[8]](#endnote-8) Similarly, if you trusted me to return the book, then, in the circumstances, returning the book constitutes satisfying your trust. Such non-causal consequences of behaviour are not “mysterious”[[9]](#endnote-9). Pettit concludes, “These observations should be enough to show … that what you do can be good in virtue of conventional or circumstantial properties, not just in virtue of causal ones; and that there is nothing particularly mysterious in taking the realization of such a property to be a constitutive as distinct from a causal consequence of your behaviour.”[[10]](#endnote-10)

*II. The Only-Behavioural-Consequences Mistake*

The good consequences discussed above are all *behavioural goods*: they depend only on the behaviour itself and the context in which it occurs. “If you keep a promise or tell the truth, it does not matter that you act out of this or that disposition; whether you act out of virtue or attachment or just sheer opportunism, you still keep the promise or tell the truth.”[[11]](#endnote-11)

Pettit argues some goods are not behavioural. They are what Pettit calls “robustly demanding”: to obtain, they require not just that things be a certain way in the scenario that contingently obtains, but that things would also be a certain way in various counterfactual scenarios. Whether your act counts as an act of friendship, honesty or respect, depends not just on whether you favour me, tell me the truth or restrain your conduct. To restrain your behaviour, tell me the truth or favour me just because it happens to suit you, when you would have acted differently if it had not been convenient, is not to give me the good associated with respect, honesty or friendship. To get at these further goods, you need to do those things robustly. When an act of favour is prompted by a robust disposition to respond to the reasons to benefit me whenever these reasons apply (absent excusing obstacles or red lights)[[12]](#endnote-12), then the act of favour constitutes as an act of friendship. This, Pettit argues, gives me an additional good over and above the act of favour.[[13]](#endnote-13) Similarly, robust restraint gives me the additional good of receiving an act of respect and robust truth telling, the additional good of receiving an act of honesty.

*III. The Only-By-Discrete-Consequences Model*

Pettit argues that it is a mistake to think that the good brought about by action can only include all-or-nothing or on-off consequences. Building on his previous argument that some goods – such as honesty, friendship and respect - require a degree of robustness, Pettit now argues that the degree of robustness with which you realise some contingent consequences can be part of the good which you bring about in action even if the degree of robustness is not high enough to realise such goods.[[14]](#endnote-14) Suppose for example, that your action springs from a disposition to favour me that is relatively robust (you would still favour me in many counterfactual scenarios) but not quite robust enough to meet the threshold to count as an act of friendship. Pettit argues that it is better for me that you act out of this relatively robust disposition than that you act out of an less robust disposition. More robust access to a good is itself good.

Pettit also argues that the threshold for robustness for goods like friendship, honesty and respect will be vague or indeterminate. The requirement in the previous discussion was idealised: respect was described as requiring that you would display restraint in all counterfactual situations where the reason for restraint remained unless there were excusing obstacles or “red lights” revealing considerations trumping respect. We do not, and should not, require this idealised robustness for someone to count as giving someone the benefit of friendship, respect, or honesty in real life. “We will be happy, presumably, to attribute a suitably robust beneficence so long as you come close to the idealised performance.”[[15]](#endnote-15) What counts as coming close enough may be indeterminate or vague. Given this lack of deteminacy, Pettit argues, we should recognise that it isn’t appropriate to describe whether you bestow goods such as friendship, honesty or respect as an all-or-nothing or on-off consequence. In addition, we should not think that providing a benefit more robustly is only better when it reaches the threshold. “If being in an appropriately marked vicinity is what is good, then it seems natural to think that steps that take you towards that vicinity must be good.”[[16]](#endnote-16)

*Models of Action: Production versus Control*

As Pettit notes, even if we recognise respect, friendship and honesty as goods over and above restraint, favour and truth, we might still be hesitant to see them as goods that could be brought about by an action. He traces such reluctance to a model of action which he calls the ‘production model’. On the production model, acting consists in producing an act: the motivating state that prompts you to act in a certain way is not part of the act, but simply a contingent cause. On this view, that very same act (with the same effects) could have been caused by a different disposition. This means that the effects of action cannot depend on the dispositions of the agent.

Pettit offers an alternative model of action, on which to act is to control for an act: to be a system that would produce that act not just in the actual circumstances but also in an appropriate range of counterfactual scenarios. Pettit uses the analogy of a temperature control system in a building: the system does not just produce a comfortable temperature in existing conditions; it would produce a temperature in that range robustly over many different variations of the weather.[[17]](#endnote-17) To understand what the system is doing, we have to think about not just its action in the actual world, but also what it would do in those counterfactual situations.

Pettit suggests that we should understand acting as controlling for an act, rather than merely producing an act. To understand what the agent is doing, we need to think not just about how the behaved in the actual world but about how they would behave in counterfactual situations. Pettit argues that this alternative model not only makes sense of the good consequences discussed in this paper, but also helps us to answer an enduring question in philosophy of action: what makes an act intentional.

Pettit’s answer to this question is that we can understand an act as intentional if and only if the production of that act was controlled for. Pettit argues that this applies both to acts that are intentional in the narrow sense – acting with the intention to achieve a certain effect – and acts that are intentional in a weaker sense – where the effect is “not desired as such but is foreseeably packaged with the desired option”[[18]](#endnote-18). He writes: “…acting intentionally, whether in the stronger or weaker senses, means controlling for the effects of the act.” Pettit thus claims that the control model is thus the only model of action that can explain why every action is intentional in the stronger sense (under some description). Pettit sees this as a powerful argument in favour of the control model.

Suppose I let go of a rope in order that a weight will drop. I control for the weight dropping. Thinking about control explains why it is not enough for my desire for the weight to drop and my belief that letting go of the rope will make this happen to have caused the action. Suppose that the onset of this belief-desire pair made me nervous, causing me to let go of the rope. I do not count as intentionally dropping the weight. On Pettit’s account, this is because I don’t meet the control condition: there are relevant counterfactual situations where the option remained within my ability and suitably appealing but I do not drop the weight.

What about action that is intentional in the weaker sense? Suppose that damaging the floor is a foreseen but not desired effect of dropping the weight. Pettit argues that I control for this effect too: “while you might prefer that the floor were not made of wood and were not exposed to damage, you still control for dropping the weight in a range of scenarios in all of which, you recognise, the floor is of that kind.”[[19]](#endnote-19) Claiming that you control for damaging the floor feels counter-intuitive. However, Pettit is using the term ‘control’ as a term of art. Your damaging of the floor is robust across at least some relevant scenarios. I think more could be said about the differences in robustness between acts that are strictly intended and those that are intended in the weaker sense. This might both reduce the intuitive discomfort with describing the weaker cases as a case of control and be fruitful as a way of understanding the differences between the two types of intention.

*The Control Model and Not-Very-Robust Actions*

Pettit begins his description of the control model using the example of giving someone respect that is already familiar from earlier discussion in the paper.[[20]](#endnote-20) If we agree that giving someone respect is an action, it does seem to be better understood along the lines of the Control Model. However, whether we should recognise giving someone respect (in the rich way Pettit understands this) as an action may be the very point at issue. There may be other actions that seem to be better understood using the Production Model. If I wiggle my finger on a whim, this seems like an intentional action in both the narrow and wider sense, and yet it may not exhibit the kind of counterfactual robustness Pettit describes.

Pettit is right that there are cases, like the respect case, where we seem to care about the robustness of a benefit across counterfactual situations. However, there are also many cases where we are not interested in what would have happened in such counterfactual situations. Pettit himself discusses some beneficial acts that do not depend on dispositions: whether you have kept your promise, told the truth, done me a favour. Although Pettit argues that any intentional action must exhibit some form of control, there is a worry here that applying the control model across the board may distort our understanding of such cases. In particular, sometimes when we say that we are interested in what someone has *done,* we mean that we want to look at their actions understood on the Production Model. We want to isolate their actions from their dispositions.

Which actions we are more interested in making sense of may not be a neutral question that is separate from the question of which ethical theory we endorse. If we care very much about respect and other similar goods, then we may be more tempted to say that whether I am giving respect rather than merely restraint makes a difference to what I *do*. We are likely to find the control model more appealing. On the other hand, if all the goods that we think are morally important are behavioural consequences of actions, then we are less likely to think that the disposition by which I act needs to be included in our identification of my action. Thus if we are looking for an argument that the production model of action should be preferred to the control model independently of our theory of the good, a lot rides on Pettit’s second argument: the claim that only the control account can explain which actions count as intention. Pettit’s argument here is promising, but I’m not yet convinced that he has shown that it is the best – let alone the only – way to answer this question.

*Closing the Gap: Convergence between Consequentialism and Non-Consequentialism*

Pettit holds that his conclusions have important implications for moral philosophy. He states that “By expanding the beneficial effects that you may bring about in action, the thesis makes it more likely that acting well is significantly related to the neutral goods that your actions bring about or promise to bring about” [[21]](#endnote-21) and “prompts us to reopen questions about the importance of neutrally good consequences and about how far the recommendations of a neutralist consequentialism may coincide with those of non-consequentialism.”[[22]](#endnote-22)

Pettit has certainly significantly narrowed the gap between neutral consequentialism and non-consequentialism. His neutral consequentialism can recognise many of the goods that non-consequentialists care about. However, the recommendations of his neutral consequentialism will still remain quite different from those of most non-consequentialists.

Pettit argues that recognising that you do me more good the more robustly you bring about a good to me – and conversely that you do me more harm the more robustly you bring about a harm to me – allows him to explain traditional doctrines like the doctrine of doing and allowing.[[23]](#endnote-23) Pettit argues that doing harm is normally worse than merely allowing harm because it normally involves controlling more robustly for its realisation. When I merely allow harm, I don’t control for harm in scenarios when others could realise it instead. However, Pettit argues, this does not vindicate the doctrine of doing and allowing in all cases. It does not vindicate a difference between doing and allowing in cases where you fail to recognise a harm to me despite being certain that no-one else will prevent it instead. In such as case, Pettit argues, your reason for not preventing the harm can’t be just general indifference, but must be indifference even in the face of certain harm to me. Given this, “you allow that harm to befall me in just as robust a manner as if you had generated it yourself”.[[24]](#endnote-24)

I have some doubts about whether doing harm does involve controlling more robustly for harm than doing harm “in most cases”.[[25]](#endnote-25) Perhaps I am less sanguine than Pettit, but I think generally when people allow harm, they would probably still allow that harm even if they *were* certain that no-one else would help instead. So I am not sure that Pettit can even vindicate as many of our intuitions about doing and allowing as he thinks he can. But even if his understanding of the implications of his account is right, the answer to his “questions…. about how far the recommendations of a neutralist consequentialism may coincide with those of non-consequentialism” is: not very far at all. This is because it is key to the doctrine of doing and allowing that allowing harm in order to avoid a given cost may be permissible even when doing harm to avoid that same cost is impermissible and everything else – including the certainty of harm – is equal. And, as I have argued elsewhere, acceptance of at least some version of the doctrine of doing and allowing is key for non-consequentialism.[[26]](#endnote-26)

Pettit mentions Peter Singer’s Pond case as an example where he does not think the doing/allowing distinction will apply. In this famous case, the agent walks past a drowning child in order to avoid getting his shows muddy. I think Pettit’s appeal to this example might be misleading, because many deontologists will agree with Singer that we are intuitively required to endure significant costs to save the child. This does not show the distinction between doing and allowing is morally unimportant when we are certain that no one else will help. In fact, most people would argue that the agent is require to save the child even if there are other people around who could also do so – so the question of certainty seems to be orthogonal to the Pond case.

The doing/allowing distinction is still intuitively morally relevant even in cases like Pond because when the cost to the agent gets high enough, we see a difference between doing and allowing. It is permissible for me to fail to pull the drowning child out of the water if I am busy saving my own life or the life of my own child; it is not permissible for me to push the child into the water so I can get past to save my life or the life of my own child. (Suppose I am trying to get to hospital and he is blocking the only route.) This is so even if we make sure that the harms and the probabilities are the same in both cases. If Pettit cannot replicate this result - and on his own statement he cannot- then the recommendations of his neutral consequentialism will be very far from those of a non-consequentialist.

*Conclusion*

“Three Mistakes about Doing Good” presents an exciting new perspective on how we should understand the good (or bad) that an action brings about. It is a paper with interesting consequences for metaphysics and moral philosophy. For metaphysics, it presents an novel way of understanding action, which may enable us to understand what it is that makes an action intentional, and even the difference between what is strictly intended and merely foreseen. In moral philosophy, it significantly narrows the gap between neutral consequentialism and non-consequentialism, allowing the neutral consequentialist to recognise many of the factors that non-consequentialists care about. However, as I have argued, Pettit does not manage to close the gap completely: his account cannot recognise key aspects of non-consequentialism such as the distinction between doing and allowing.

1. I’ll focus on the good henceforth. [↑](#endnote-ref-1)
2. Philip Pettit, “Three mistakes about doing good (and bad)”, *Journal of Applied Philosophy*, p. 1. [↑](#endnote-ref-2)
3. Pettit op. cit., p. 21 [↑](#endnote-ref-3)
4. Pettit op. cit., p. 3 [↑](#endnote-ref-4)
5. Pettit op. cit., p. 37 [↑](#endnote-ref-5)
6. Pettit op. cit., p. 11-12 [↑](#endnote-ref-6)
7. Pettit op. cit., p. 12. Here Pettit cites Alvin Goldman, *A Theory of Human Action*, (Englewood-Cliffs, NJ: Prentice Hall, 1970). [↑](#endnote-ref-7)
8. Pettit op. cit., p. 13. [↑](#endnote-ref-8)
9. Pettit op. cit., p. 13 [↑](#endnote-ref-9)
10. Pettit op. cit., p. 14. [↑](#endnote-ref-10)
11. Pettit op. cit., p. 15. [↑](#endnote-ref-11)
12. Pettit op. cit., p. 18. [↑](#endnote-ref-12)
13. Pettit op. cit., p. 19 [↑](#endnote-ref-13)
14. Pettit op. cit., p. 24 [↑](#endnote-ref-14)
15. Pettit op. cit., p. 26. [↑](#endnote-ref-15)
16. Pettit op. cit., p. 26. [↑](#endnote-ref-16)
17. Pettit op. cit., p. 31. [↑](#endnote-ref-17)
18. Pettit op. cit., p. 34 [↑](#endnote-ref-18)
19. Pettit op. cit., p. 30 [↑](#endnote-ref-19)
20. Pettit op. cit., p. 34 [↑](#endnote-ref-20)
21. Pettit op. cit., p. 3. [↑](#endnote-ref-21)
22. Pettit op. cit., p. 37 [↑](#endnote-ref-22)
23. Pettit op. cit., p.28, see also Philip Pettit, *The Robust Demands of the Good: Ethics with Attachment, Virtue and Respect* (Oxford: Oxford University Press, 2015), Ch. 6. [↑](#endnote-ref-23)
24. Pettit op. cit., p. 191. [↑](#endnote-ref-24)
25. Pettit op. cit., p. 28. [↑](#endnote-ref-25)
26. Fiona Woollard, ‘Double effect, doing and allowing, and the relaxed nonconsequentialist’, *Philosophical Explorations*, 20:sup2 (2017), 142-158. [↑](#endnote-ref-26)