

Fitting Attitudes, Reasons for Action, and the Rejection of Consequentialism

John Ku
Howard Nye
David Plunkett

Abstract:

We argue that consequentialism may seem attractive, despite the well-known case intuitions against it, primarily due to the idea that it is clear how we have reason to bring about good states of affairs but entirely mysterious how we could have reason to do anything else. We show that a deep theoretical connection between good states of affairs and reasons to bring them about can indeed be vindicated by the conceptual connections between good states of affairs, fitting desires that they obtain, and reasons to act out of them. We argue, however, that the conceptual connection between fitting attitudes and reasons for action equally vindicates reasons to do things other than bring about the best states of affairs. The fittingness of motives other than desires for good states of affairs – some more intimately related to morality than such desires – undermines consequentialism whether as a theory of rationality or morality.

I. Why Consequentialism?

Consequentialism is the view that what an agent has most reason or most moral reason to do is determined solely by the goodness of some resulting states of affairs. This may take the form of act consequentialism, according to which one's (moral) reasons to perform an act are determined directly by whether the act brings about the best consequences. Or it may take some form of indirect consequentialism; for instance, rule consequentialism maintains that one's (moral) reasons to perform an act are determined by whether the act conforms to rules which, if followed or accepted, would bring about the best consequences.

In this paper, we will argue that the central theoretical motivation behind consequentialism is mistaken. Without this motivation to support it, we think that the well-known case intuitions against consequentialism are decisive. We will argue explicitly against act consequentialism, but we think that our arguments can generalize to indirect consequentialism as well. In order to introduce the central theoretical motivation for consequentialism, consider the

following traditional dilemma faced by consequentialists:

Horn 1: Act Consequentialism faces “death by a thousand counterexamples.”¹

Act consequentialism delivers many results that are badly in conflict with our intuitions about what we have most (moral) reason to do in specific situations. As one example, suppose a doctor is considering secretly killing and cutting up one healthy person in order to transplant his organs to save the lives of five others. On the uncontroversial assumption that the five people dying is worse than the one dying, act consequentialism delivers the strongly counterintuitive result that it is not only morally permissible or rational to kill the one to save the five, but in fact that it would be morally wrong or irrational not to.

Some consequentialists may try to deny that this result really follows from consequentialism. They might hold, for instance, that the killing of an innocent person is so intrinsically bad as to outweigh the badness of four additional deaths. Apart from worries about the plausibility of such a theory of goodness, this defense merely causes strongly counterintuitive results to re-arise in slightly altered cases. We need only consider cases in which a doctor can prevent two such killings by performing one himself. Even counting such killings as intrinsically bad, act consequentialism is again committed to the result that it is not only morally permissible or rational for the doctor to kill and cut up the healthy person, but in fact morally wrong or irrational not to.

This is by no means the only case where consequentialism yields counterintuitive results. Because it looks only to the consequences of an action, act consequentialism gives no intrinsic

¹ Thanks to David Braddon-Mitchell (2007) for this way of phrasing the problem.

weight to keeping promises or to punishing only the guilty.² And since one's personal relations and projects only matter insofar as they contribute to the overall goodness of a state of affairs, it also runs counter to our intuitions that there are important reasons and special obligations to the near and dear as well as intuitions concerning how much one can be demanded to sacrifice for the sake of the total good. Given these difficulties accommodating our case intuitions, one might justifiably start to wonder what ever attracted philosophers to act consequentialism to begin with.

Horn 2: Indirect consequentialism lacks motivation.

Given the difficulty act consequentialism faces, one might think moving to some form of indirect consequentialism would avoid delivering such intuitively unpalatable results. Whether or not it in fact would, our present concern is with why so many philosophers have felt dissatisfied with indirect consequentialism, seemingly regardless of its potential to better capture our case intuitions. J.J.C. Smart's (1967, 1973) accusation of rule consequentialists as guilty of an irrational form of “rule worship” is here quite revealing. It seems we can legitimately demand an answer to the question, “Why should I follow those rules which would have the best consequences if everyone followed them?” After all, people are not in fact following those rules and my own act of following such rules is not going to change that.

What is revealing is that Smart does not similarly accuse the act consequentialist of “act-”

² See Darwall (1998, 131-132) for this diagnosis of these kinds of counter-intuitive results in terms of the lack of intrinsic weighting of considerations like breaking promises and punishing the innocent. Even consequentialist theories that hold that these kinds of features contribute intrinsically to the badness of states of affairs that involve them still do not allow these features themselves to count against performing such actions independent of considerations of the goodness or badness of the states of affairs they bring about, which conflicts with our intuitions to the effect that they can (as can be seen in the kinds of counter-intuitive results that will re-arise in a way analogous to that of performing one organ cutting to prevent two others – e.g. from breaking one promise to prevent the breaking of two others, etc.).

or “goodness worship,” or think we need a substantive answer to the question, “Why act in ways that will bring about the best consequences?” The theoretical motivation that act consequentialism has, and indirect consequentialism can be seen to have strayed from, thus appears to be the following:

Basic Theoretical Motivation for Consequentialism:

It is clear how good states of affairs give us (moral) reasons for action, but entirely mysterious how any other considerations can give us (moral) reasons for action.

This motivation would clearly support consequentialism as a global theory of our reasons for action, but it is also fairly easy to see how it might support consequentialism as a moral theory. Consequentialists like Sidgwick (1874, 1889) might maintain that there are other unmysterious sources of reasons, such as avoiding our own pain, but that the only remotely moral-looking reasons are those to bring about the overall best state of affairs.

We think that this basic theoretical motivation also accounts for why act consequentialists think that their theory can survive its counter-intuitive predictions in so many cases. Act consequentialism is held to enjoy a source of theoretical support not dependent on fitting our case intuitions; the unparalleled clarity of our reasons to bring about good states of affairs seems to be evidence in favor of it. Moreover, if correct the basic theoretical motivation would give us reason in advance to think that any intuitions that run contrary to consequentialism must be somehow mistaken and debunkable, suggesting that consequentialism is, after all, the theory that best unifies and explains our intuitions on reflection.

But in light of how much work the basic theoretical motivation is doing here, we think

that without its support, act consequentialism is unable either to ignore or so easily to debunk intuitive counter-evidence. We thus think that consequentialism stands or falls with its basic theoretical motivation.

We think that there is indeed an important truth behind the basic theoretical motivation for consequentialism. We will argue that it is a conceptual truth that if a state of affairs is good, we have reason to bring it about, thus securing a clear connection between good states of affairs and reasons to bring them about. However, we will also argue that once one understands why this is a conceptual truth, one should see that it can be just as clear and unmysterious how we can have reason to do many other kinds of things, thereby undermining the basic theoretical motivation for consequentialism.

II. Good States of Affairs and Fitting Attitudes

Let us begin our explanation of the conceptual connection between good states of affairs and reasons for action by starting with the concept of a GOOD STATE OF AFFAIRS. We think that the analysis that best captures the content and normative force of claims about which outcomes are good is the following:

Fitting Attitude Analysis of Good States of Affairs:

State of affairs S is good iff we have fittingness reason to agent-neutrally desire that S obtains.³

³ See e.g. Ewing (1939), Gibbard (1990), and Scanlon (1998)

The fittingness reasons (aka. “reasons of the right kind”) spoken of here are to be contrasted with pragmatic or strategic reasons to get oneself to have a desire. To illustrate the distinction, suppose that an evil demon threatened to harm your loved ones if you do not desire that there be an odd number of hairs on your head. You would then have strong pragmatic reasons to get yourself to have such a desire. However, this would not thereby make it good that you have an odd number of hairs on your head. Given the circumstances, it might be good to have the desire, but it would not make the object of the desire good. It is characteristic of pragmatic reasons to have an attitude that judgments that one has such reasons are not capable of directly guiding one's coming to have the attitude. Coming to have an attitude in response to apparent pragmatic reasons requires further goal-directed behavior or activity directed towards getting oneself to have it, such as classically conditioning oneself, taking mind-altering substances, or paying selective attention to features of the situation, etc. Judgments that an attitude is fitting or appropriate, however, are capable of directly guiding one's having it without any intermediation by goal-directed behavior to get oneself to have it. For instance, judging that it is fitting or appropriate to desire knowledge for its own sake - as one might conclude upon contemplating Nozick's (1974) experience machine - can directly cause one to desire knowledge intrinsically. It is these fittingness reasons that we argue figure into the analysis of good states of affairs.

Moreover, the desires in question should be understood as those that are *agent-neutral*, or desires that a state of affairs obtains, as opposed to those that are *agent-relative*, or desires to directly do something or bring about a state of affairs oneself. To see why it is only agent-neutral desires that are relevant here, consider the feeling of obligation one might have not to step on

someone's foot.⁴ One might judge that this feeling of obligation is fitting and be more motivated not to step on her foot than to prevent her having her foot stepped on by someone else. However, this surely would not necessarily involve thinking that one's stepping on her foot would be a worse state of affairs than her having her foot stepped on by someone else. The motivation here is to *not step* on her foot rather than to merely bring it about her foot is not stepped on, but reasons to have motivations of this kind do not contribute to the goodness or badness of the state of affairs that they motivate one to bring about.

In order to see the strength of this fitting-attitude analysis, consider the various states of affairs people have intelligibly held to be good. These would include not only such things as happiness, knowledge, beauty, and achievement, but also all manner of apparently miscellaneous things such as abstinence, racial purity, tradition, obedience to deities, etc. It seems that nothing unifies the content of all these judgments except that these are all states of affairs we are judged to have reason to desire.

Moreover, this analysis seems to capture the normative force that judgments of goodness have in directly guiding desires. Even if someone were to use the word 'good' to label all those things we would call good, but did not take that to guide desires, it seems that by 'good', she wouldn't really mean good. On the other hand, if someone were to label as 'good' precisely those things we think are bad, she would still be perfectly intelligible as thinking those things are good so long as she takes that to mean that we have reason to desire them.

⁴ This example is from Darwall (2006), though he does not describe one's response in such a circumstance in terms of feelings of obligation.

III. Fitting Attitudes and Reasons for Action

It thus appears that the states of affairs which are good are those which we have fittingness reason to agent neutrally desire. With the addition of the following intuitively compelling conceptual principles, we think that this conceptual analysis can secure the clear connection between good states of affairs and reasons to bring them about that motivates consequentialism. The first is:

Warrant composition principle [WCP]:

Let ϕ be a psychic state that involves ψ as an essential component. If we have fittingness reason to be in ϕ , then we have fittingness reason to be in ψ .

WCP simply states that if one has reason to be in a psychic state, one necessarily has reason to be in all that that psychic state necessarily involves. In particular, it seems to be a conceptual truth that if one has a desire, one necessarily has motivation to bring about the state of affairs that is the object of the desire. If one lacks that motivation, one simply doesn't count as having the desire. Thus, WCP entails that if one has reason to desire some state of affairs, one has reason to be motivated to bring it about. Our second principle is:

Motivations-actions principle [MAP]:

Let ϕ be an action. If one has fittingness reason to be motivated to ϕ , then one has reason to actually ϕ .⁵

⁵ WCP and MAP closely correspond to Skorupski's partition of his "Feeling/Disposition Principle," (FD): "if there is reason to feel ϕ then there's reason to do that which ϕ disposes to" into (FDF): "If there's reason to feel ϕ there's

The idea behind MAP is that since it is through the rational assessment of our motivations that our actions are guided by reason, our reasons for action just are reasons for motivations to perform these actions.⁶ Applying this principle, we can see that reasons to be motivated to bring about a state of affairs would necessarily entail reasons to actually bring it about.

The explanation for the clear connection between good states of affairs and reasons to bring them about thus runs as follows. If a state of affairs is good, we have reason to agent neutrally desire that it obtains. Since this desire necessarily involves motivation to bring about the state of affairs, we have reason to be motivated to bring it about. Reasons to be motivated to bring about that state of affairs in turn entail that we have reasons to actually bring it about. Hence, if a state of affairs is good, we have reason to bring it about.

IV. The Theoretical Motivation Revisited

Recall that the basic theoretical motivation for consequentialism was not just the view that there is this kind of clear connection between good states of affairs and reasons to bring them about, but also that it is mysterious how any other considerations could give us reasons for action.

However, given our vindication of the clear connection between goodness and reasons for action in terms of a general conceptual connection between fitting motives and reasons to act out of

reason to desire to do that which ϕ characteristically disposes one to desire to do,” and (FDD): “If there’s reason to desire to do α (or to bring it about that p), there’s reason to α (to do that which will bring it about that p)” (see Skorupski (1999), especially p. 38, 63, 131, and 174 n24). The main possible difference between our principles and Skorupski’s FD principles is our WCP’s insistence on ψ ’s essential involvement as component of ϕ rather than (in his FDF) its mere involvement as what “ ϕ characteristically disposes one to desire to do,” and our MAP’s clarification that the relationship between reasons for motivations and actions is held to hold only between fittingness (and not pragmatic) reasons for motivation and reasons to act out of them (as well, perhaps, as our MAP’s clarification that this relationship holds between all fitting motives and reasons to act out of them, not just those motives that are ‘desires’ in the ordinary English sense – though we think it clear from context that Skorupski intended both of these features our MAP makes explicit).

⁶ Remarks by Scanlon (1998), Gibbard (1990) and Bratman (1987) are suggestive of this idea.

them, we think it should be evident that the very explanation of this clear connection can equally ground non-consequentialist reasons for action. Agent-neutral desires are, after all, just one out of many different types of human attitudes. WCP and MAP assure a conceptual connection between *any* fitting motivationally laden attitudes and reasons to act out of them. There are many motivationally laden attitudes that are not agent-neutral in character, including feelings of obligation, guilt, anger, shame, scorn and the sense of honour (just to name a few). These states involve motivations that are agent-relative in character.

For instance, as we saw above, feelings of obligation (i.e. a forward-looking kind of guilt tinged motivation) involve a motivation to refrain from performing certain acts. The motivational content of this attitude is not adequately captured by an agent-neutral desire that a state of affairs be brought about in which performances of the act in question are minimized.⁷ When these attitudes are fitting, then, there will be just as clear and non-mysterious a connection to agent-relative or non-consequentialist reasons for action as there is between fitting agent-neutral desires and reasons to bring about good states of affairs. This is ensured by the same two principles that ground the conceptual connection between goodness and reasons for action. Since feelings of obligation necessarily involve a direct motivation not to perform the act in question,

⁷ Nor is it adequately captured by an agent-neutral desire that a state of affairs be brought about in which performances of the act in question by oneself in particular are minimized. For instance, consider the following variation of the classic organ cutting case of J.J. Thomson (1985):

Suppose that what had happened was this: The surgeon was financially badly overextended last fall, he had known he was named a beneficiary in his five patients' wills, and it swept over him one day to give them chemical X to kill them ["Now chemical X works differently in different people. In some it causes lung failure, in others kidney failure, in others heart failure. So these five patients who now need parts" need them because the surgeon intentionally gave them the chemical to kill them for their inheritances]. Now he repents, and would save them if he could. If he does not save them, he will positively have murdered them. Does *that* fact make it permissible for him to cut the young man up and distribute his parts to the five who need them? (Thomson (1985), 98).

The intuitive answer, as Thomson points out, is clearly that it is still quite impermissible to kill the one to save (or here prevent one's killing) the five. But all of us who get this intuition likely have (and intuitively take to be fitting) feelings of obligation not to kill – not merely to bring it about that one minimizes the killing one does!

WCP entails that reasons to have this attitude result in reasons to be motivated not to perform the act in question. MAP then implies that reasons to be motivated not to perform the act result in reasons not to perform the act. Similar connections can be drawn between the fittingness of many other attitudes and non-consequentialist reasons.

Since the vindication of the connection between good states of affairs and reasons to bring them about thus equally vindicates the connection between the fittingness of other, non-agent-neutral attitudes and reasons for action, we conclude that the basic theoretical motivation for consequentialism is entirely undermined. One could still coherently maintain the truth of consequentialism, but in doing so we think that one would be committed to rather bizarre entailments. There are two ways in which consequentialism could still be true for all we have said: (i) no motives other than agent-neutral desires are ever fitting, or (ii) the motives involved in other fitting attitudes perfectly align with those of fitting agent-neutral desires.

But why think that either of these is remotely plausible? Why should agent-neutral desires be so privileged over other kinds of motives? We suspect that only a conflation between fittingness and pragmatic reasons could motivate this prejudice. One might think, for instance, that guilt is just a negative, destructive emotion that it is not good to feel or perhaps that it should only be felt in those cases where it is desirable to feel it. But this is precisely to conflate pragmatic reasons to feel guilt with fittingness reasons. Suppose that out of envy towards a friend you intentionally deceived a potential employer about his job qualifications, costing him his dream-job. You feel guilty for having done this, but realize that doing so will decrease your productivity slightly – making it the case that you will have less money to donate to Oxfam. Not feeling guilt might thus bring about a better overall state of affairs, but this would in no way make guilt any less fitting or appropriate.

To better see the folly of neglecting fittingness reasons, consider the analogous conflation in the case of reasons to have agent-neutral desires. If an evil demon threatens to torture one's loved ones unless one desires their suffering, then it would be good to have the desire for their suffering. But it certainly doesn't make it fitting to desire their suffering and therefore good for them to suffer. These are merely pragmatic reasons to get oneself to have such a desire. It would be circular for a consequentialist to attempt to evaluate the rationality of all attitudes *including* agent-neutral desires solely in terms of whether or not they bring about good consequences – good consequences must themselves be understood in terms of fittingness reasons for agent-neutral desires. But it would be absurdly unprincipled to evaluate only the rationality of agent-neutral desires, and not other attitudes, in terms of fittingness rather than pragmatic reasons.

V. Decision Theoretic Interlude

Recognizing genuine fittingness reasons for attitudes other than agent-neutral desires will thus undermine the theoretical motivation for consequentialism. Consequentialism about rationality would be refuted directly by the existence of the other reasons for action they generate. We should emphasize, however, that rejecting consequentialism as a theory of rational action for these reasons in no way creates problems for any suitably interpreted decision theory – especially since so many seem to think otherwise. Decision theory is best understood as a theory of relative rationality – that is as a theory of what one should do given a rational objective function specifying one's rational ends or non-instrumental motivations.⁸ Given the understanding of the

⁸ See Darwall (1983) and Gibbard (1998). We think that things here are analogous to the appropriate interpretation of Bayesian theories of credence updating in epistemology – they are theories of relative rationality as well, namely theories of what credences to have given that one has rational prior credences (where something like a procedure of inference to the best explanation is needed to determine as a matter of substantive normative fact what the rational priors are) – see Gibbard (1998) on this point as well.

relationship between reasons for motivationally laden attitudes and reasons for action afforded by the warrant composition and motivations-actions principles, we can see that the rational objective function is determined via the determination of which motivationally laden attitudes are fitting. Decision theory is (or at least should be) completely non-committal on this rational objective function's being agent neutral in character; the rational objective function specifies which states of affairs to prefer, but it is entirely left open as to whether or not these states of affairs are specified in agent-relative terms that make essential indexical reference to the agent performing the acts in question at the time of decision⁹ (e.g. as would be involved in a preference out of fitting feelings of obligation to refrain from stepping on someone else's foot). Our arguments against consequentialism as a theory of rational action are thus best understood as arguments that the rational objective function will indeed involve preferences over states of affairs that make such essential indexical reference.

VI. Morality and Reasons for Action

We have thus argued that the basic theoretical motivation for consequentialism fails to successfully motivate it as a theory of reasons for action generally. But one might wonder if consequentialism could still retain a kind of theoretical motivation as a theory of our moral reasons for action in particular. One might concede that there are fitting attitudes that translate into reasons for action other than agent-neutral desires, but still maintain (as in Section I we suggested one might) that our only *moral*-looking reasons are our reasons to bring about good states of affairs. The thought might be that only our agent neutral-reasons¹⁰ are sufficiently

⁹ Cf. Hurley (1997).

¹⁰ I.e. our reasons to act out of the fitting agent neutral- desires that constitute the goodness of states of affairs.

selfless, disinterested, or impartial to be our moral reasons. We think that a closer look at what our moral reasons are can dispel this theoretical motivation for consequentialism as moral theory in particular.

For reasons very similar to those in support of our fitting attitude analysis of GOOD STATES OF AFFAIRS, we think that the following analysis best captures the content and normative force of claims about moral wrongness:

The Quasi-Gibbard Analysis of Moral Wrongness:¹¹

Agent A's act of ϕ -ing is morally wrong if and only if it is fitting for A to feel obligated not to ϕ (or equivalently fitting for A to feel prospective guilt-tinged aversion towards ϕ -ing¹²).

Consider the wide diversity of phenomena that people have intelligibly thought to be morally wrong: inflicting harms, defecting in collective action problems, failing to respect autonomy, but also all manner of apparently miscellaneous behavior, including sexual practices, drug use,

¹¹ We say "quasi-Gibbard" because Gibbard himself does not give this analysis explicitly by analyzing moral wrongness in terms of warrant for prospective guilt-tinged aversion. This analysis pretends to the title of "quasi-Gibbard", however, because it seeks to analyze a moral concept in terms of the rationality of moral emotions, and we think was what Gibbard was trying to get at in his explicit account of moral wrongness as given in Gibbard (1990). Gibbard (2006, 2007) seems to take a favorable attitude towards this analysis, or at least the use of prospective guilt-tinged aversion in the analysis of moral concepts.

¹² Prospective guilt-tinged aversion, aka. feelings of obligation, aka. motivation out of a sense of duty, is the attitude that one characteristically feels upon contemplating the prospect of performing or failing to perform an action, which act or omission one thinks would constitute doing something morally wrong. It is discussed admirably by Brandt (1959) (see especially his discussion of the "parked car episode" p.117-118), and was what Mill (1863) described as an "internal sanction of duty... a feeling in our own mind... attendant on violation of duty, which in properly cultivated moral natures rises, in the more serious cases, into shrinking from it as an impossibility," and "a mass of feeling which must be broken through in order to do what violates our standard of right." Some might contend that rather than simply an emotion this attitude must be understood as a subconscious judgment of moral obligation or wrongness, but it seems that it can be felt without any such judgment (or even sub-judgmental appearance to this effect) and that it is best understood as motivational syndrome in a way that any account of it as a judgment could not capture (though a full defense of this anti-judgmentalist order of explanation is beyond the scope of this paper).

swearing, “playing God,” etc. It seems that the only thing that unifies the content of all of these judgments is that those who make them think that one should feel obligated not to perform such acts.

Similarly, just as judgments of the goodness of states of affairs have the central normative feature of guiding attitudes (in particular agent-neutral desires), moral judgments seem to have this central normative property as well – in particular by guiding feelings of prospective guilt-tinged aversion. Even if someone were to label as ‘morally wrong’ all those things we would call wrong, but did not take that to guide feelings of obligation, it seems that by ‘wrong’, she wouldn’t really mean wrong. On the other hand, if someone were to label ‘morally wrong’ precisely those things we think permissible, she would still be perfectly intelligible as thinking that those things are wrong so long as she takes this to mean that one should feel obligated not to perform them.

Hence, consequentialism about morality amounts to the thesis that the only thing we have reason to feel obligated not to do is failing to bring about the best states of affairs. If one thought that there was a privileged connection between good states of affairs and reasons for action, then this thesis might make sense. Morality seems to give us reasons for action, so if good states of affairs are the only reason giving – and as opposed for instance to self interest the only remotely moral-looking – game in town, morality presumably prescribes only that we bring about good states of affairs. But if, as we have argued, good states of affairs have no privileged connection to reasons for action, the consequentialist thesis is still coherent, but this motivation is entirely undermined.

Indeed, we can see that the exact same considerations that vindicated a conceptual connection between good states of affairs and reasons to bring them about equally vindicate an

immediate conceptual connection between the moral wrongness of an action and reasons not to perform it. There is no need for any intermediation by the link between goodness and reasons for action. According to the quasi-Gibbard account of moral wrongness, if an action is wrong, one has reason to feel obligated not to perform it. But feeling obligated not to perform it involves being motivated not to perform it. Thus, by the warrant composition principle if one has reason to feel obligated not to perform the action, one has reason to be motivated not to perform it. But by the motivations-actions principle if one has reason to be motivated not to perform an action, one has reason not to perform it. Hence, if an action is morally wrong one has direct reason not to perform it.

VII. Conclusion

We thus think the basic theoretical motivation for consequentialism fails, leaving the consequentialist with no adequate response to the large and compelling body of intuitive counter-evidence. In light of this, and our having established a clear and non-mysterious way for there to be reasons to act out of motives other than desires for good states of affairs, we conclude that consequentialism as a theory of rationality must be rejected. Moreover, since some of these motives are even more intimately connected to morality than desires for the good, we have established a clear and non-mysterious way for there to be non-consequentialist moral reasons for action. Because of this, we conclude that the intuitive counter-evidence is also decisive against consequentialism as a theory of moral reasons in particular, and that it too must be rejected.

REFERENCES

- Braddon-Mitchell, David. 2007. "Freedom and Binding Consequentialism" Presentation, University of Michigan.
- Brandt, Richard B. 1959. *Ethical Theory*. Englewood Cliffs, N.J: Prentice Hall.
- Bratman, Michael E. 1987. *Intention, Plans, and Practical Reason*. Cambridge, Massachusetts: Harvard University Press.
- Darwall, Stephen L. 1983. *Impartial Reason*. Ithaca: Cornell University Press, 1983.
- Darwall, Stephen L. 1998. *Philosophical Ethics*. Boulder, CO: Westview Press.
- Darwall, Stephen L. 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge: Harvard University Press.
- Ewing, A.C. 1939. "A Suggested Non-Naturalistic Analysis of Good." *Mind*, 48: 1-22.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge, Massachusetts: Harvard University Press.
- Gibbard, Allan. 1998. "Preference and Preferability." In Christoph Fehige and Ulla Wessels, eds., *Preferences*. Berlin: Walter de Gruyter.
- Gibbard, Allan. 2006. Reply to Critics. *Philosophy and Phenomenological Research*. 72: 729 - 744.
- Gibbard, Allan. 2007. "Thinking How to Live Together." In Grethe B. Peterson, ed., *The Tanner Lectures on Human Values*, vol. 27 (Salt Lake City: University of Utah Press), 165–226.
- Hurley, Paul. 1997. "Agent-Centered Restrictions: Clearing the Air of Paradox." *Ethics*, 108(1): 120-146.
- Mill, John Stuart. 1863. *Utilitarianism*. London: Parker, Son, and Bourn.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge, Massachusetts: Harvard University Press.
- Sidgwick, Henry. 1874. *The Methods of Ethics*, London: Macmillan.
- Sidgwick, Henry. 1889. "Some Fundamental Ethical Controversies", *Mind* 14: 473-87; reprinted in Singer, ed., 2000.

Skorupski, John. 1999. *Ethical Explorations*. New York: Oxford University Press.

Smart, J.J. C. 1967. "Extreme and Restricted Utilitarianism." In *Theories of Ethics*, ed. Philippa Foot. Oxford: Oxford University Press. P.171-183.

Smart, J.J. C. 1973. "An Outline of a System of Utilitarian Ethics." In *Utilitarianism: For and Against*, by J.J.C. Smart and B. Williams. Cambridge: Cambridge University Press. p. 3 – 74.

Thomson, Judith Jarvis. 1985. "The Trolley Problem." *Yale Law Journal*, 94: 1395-1415.
Reprinted in J.J. Thomson and William Parent, eds., *Rights, Restitution, and Risk: Essays in Moral Theory*. Cambridge: Harvard University Press, 1986, p.94-116.