Appendix: Hodgson's Argument

Introduction

In *Consequences of Utilitarianism,* D. H. Hodgson tries to show that act utilitarianism is self-defeating (in a sense to be explained later).[1] By "act utilitarianism" Hodgson understands the following principle:

> (AU) An act is right if and only if it would have best consequences, that is, consequences at least as good as those of any alternative act open to the agent.[2]

By "consequences" are here understood actual consequences, but, says Hodgson, his arguments also apply to principles which refer to probable or reasonably foreseeable or foreseen consequences.[3] All such principles, Hodgson claims, are self-defeating; that is, he claims that

> even correct application of act-utilitarianism, either by everyone in a community [where there is common knowledge that everyone is an act-utilitarian], or by individuals in a non-act-utilitarian society like our own, would not necessarily have better consequences, and

[1] D. H. Hodgson, *Consequences of Utilitarianism: A Study in Normative Ethics and Legal Theory* (Oxford: Oxford University Press, 1967). (Hodgson himself does not use the term "self-defeating".)

[2] Ibid., p. 1. Hodgson does not—like, e.g., Moore—distinguish between obligatory and merely right actions, and he apparently takes it for granted that an action is wrong if, and only if, it is not right.

[3] Ibid., p. 13. According to Allan Gibbard, Hodgson's arguments apply *only* to such principles: "Now theories of what is objectively right have no direct bearing on the problem Hodgson raises. Hodgson's thesis concerns the behaviour of rational act-utilitarians who do not know what to expect from each other: they know they lack relevant information. In order to know what they will do, we need to know how they base their decisions on information they know is incomplete. (Allan Gibbard, "Act-Utilitarian Agreements", in A. I. Goldman and J. Kim (eds.), *Values and Morals* (Dordrecht: Reidel, 1978): 91–119; p. 96.) As far as I can see, however, all this is consistent with taking AU as a theory of objective rightness.

J. Österberg, *Towards Reunion in Ethics*, Philosophical Studies Series 138,
https://doi.org/10.1007/978-3-030-12410-6

> would probably have worse consequences, than would acceptance of specific conventional moral rules and personal rules.[4]

In both cases, Hodgson says, AU is in dire straits with respect to two very useful "practices" ("institutions"), viz. those of promising and truth-telling. According to AU, keeping a promise and telling the truth do not, *per se*, constitute reasons for acting, so these practices will be absent in a society where (there is common knowledge that) everyone accepts and conforms to AU. This, argues Hodgson, gives the utilitarian agent a severe handicap in his endeavour to make the world as good as possible, for the absence of these practices in a society means that co-operation is absent too. And the absence of co-operation makes the act-utilitarian society (henceforth the AU-society) totally different from all existing societies, depriving it of most of the fruits of civilization: "there could be no human relationships as we know them".[5] And this would, of course, deprive the AU-society of a great amount of value. But, as Hodgson also stresses, this is fully compatible with everyone's fully conforming to AU.

> For this means only that the consequences would be the best possible in the circumstances; and since the circumstances (universal acceptance and rational application of act-utilitarianism, and common knowledge of this) preclude human relationships, the best possible consequences in these circumstances would be worse than consequences which are not the best possible in other more favourable circumstances.[6]

Likewise the consequences of being an act-utilitarian in a non-AU-society would be worse than those of being an adherent of common-sense morality (CSM).

If Hodgson is right, AU is, to use Parfit's notion, *directly collectively* self-defeating for people in the AU-society. (In what sense it is self-defeating for act-utilitarians in non-AU-societies will be considered in section "Truth-telling and promise-keeping in non-AU-societies" below.) A theory *T* is directly collectively self-defeating, Parfit says, when

> [i]t is certain that, if we [the members of some group] all successfully follow T, we will thereby cause our T-given aims to be worse achieved than they would have been if none of us had successfully followed T […].[7]

---

[4] Ibid., p. 38. This is a very guarded formulation of the claim. To accord with the general tenor of Hodgson's exposition, "probably" should be replaced by "almost certainly" or something of that sort.

[5] Ibid., p. 45. It would not, of course, be a society in the proper sense of the word.

[6] Ibid.

[7] Derek Parfit, *Reasons and Persons*, p. 54. According to Parfit, AU, being a species of consequentialism (C), *cannot* be directly collectively self-defeating. For, Parfit argues, "[w]e successfully follow C when each does the act which, of the acts that are possible for him, makes the outcome best. If our acts do jointly produce the best outcome, we must all be successfully following C″ (ibid.). Now Hodgson does not deny that—given the circumstances, viz. universal acceptance and rational application of act utilitarianism, and common knowledge of this—people in the AU-society always jointly produce the best outcome and that, this being the case, each of them is successfully following AU. What he claims is that, if they had *instead* accepted and rationally followed a certain other morality, roughly CSM, their circumstances then being different, they would, on the whole, have jointly produced better outcomes than the ones that they did produce. And, if this is so, then

Hodgson's arguments were hotly disputed by several prominent moral philosophers. Hodgson himself did not respond to the criticism, and I know of only three published defences of his theses.[8] The general impression was, no doubt, that Hodgson was wrong and had lost his case. In my opinion, however, Hodgson was (mainly) right and his critics were (mainly) mistaken. I will therefore reopen the case and try to prove Hodgson right, not so much by adding new arguments for Hodgson's claims as by criticizing the objections brought against it. I will begin with the case of AU in the AU-society and the fate of truth-telling in this society (Sections "Truth-telling in the AU-society: Hodgson's argument" and "Truth-telling in the AU society: objections"). I will then discuss the fate of promise-keeping in the AU-society (Sections "Promise-keeping in the AU society: Hodgson's argument" and "Promise-keeping in the AU society: objections") and, rather briefly, the case of being an act-utilitarian in a non-AU-society (Section "Truth-telling and promise-keeping in non-AU-societies"). The Appendix ends with some concluding remarks (Section "Concluding remarks").

Truth-Telling in the AU-Society: Hodgson's Argument

Hodgson asks us to consider a society in which everyone (i) accepts AU as his only personal rule; (ii) always tries to act in accordance with it; (iii) is highly rational and understands all relevant implications of the previous two conditions; and (iv) knows of the previous three conditions, knows that everyone else knows of them, and so on. In this society, Hodgson says, there would be no communication. In actual non-AU-societies, such as ours, there are several good reasons for a person to think that what other people tell him is taken to be true by them; chief among these is the fact that truth-telling is required by CSM, which most people adhere to. In actual societies, therefore, telling the truth usually has better consequences than lying. For if truth-telling is required by their morality, members of a non-act-utilitarian society will ordinarily rely on what they are told; and because of that they will often make arrangements, based on what they are told, which will have bad consequences unless they are told the truth. Also, people will resent those who lie to them; they will blame them and count less on them in the future, all of which are harmful consequences of acts of lying.

(*pace* Parfit) AU obviously *can* be directly collectively self-defeating. (Perhaps, however, the opposition between Parfit and me is due to our understanding the self-defeatingness condition differently.)

[8] See Adrian Piper, "Utility, Publicity, and Manipulation", *Ethics*, 88 (1978): 189–206 (discussed in Subsec. 3: (iv) below); Donald Regan, *Utilitarianism and Cooperation* (Oxford: Oxford University Press, 1984) (mentioned in Subsecs. 3: (ii) and (iii) below); and C. Provis, "Gauthier on Coordination", *Dialogue: Canadian Philosophical Review*, 16 (1977): 507–9 (discussed in Subsec. 3: (i) below). In addition, Hodgson is shortly mentioned with approval in G. J. Warnock, *The Object of Morality* (London: Methuen, 1971), p. 33 f., and in Dan Brock, Recent Work in Utilitarianism", *American Philosophical Quarterly*, 10 (1973): 241–76; p. 258.

In the AU-society, however, *A* has a reason to tell *B* the truth (rather than what is false) only if he has reason to believe that *B* will take what he is told as true (rather than false). But, as *A* knows, *B* will take what he is told as true only if he has reason to believe that *A* has a reason to tell him the truth—which completes the circle. Obviously, the same circularity is involved in *A's* attempt to communicate the truth by telling *B* what is false. Hodgson therefore concludes that in the AU-society no one would take what he is told as more likely to be true than false, or vice versa; attempts to communicate information would therefore be pointless.

The central passage of Hodgson's argument runs as follows:

> Being highly rational, the informant would know that the taking of the information as true rather than false was a condition precedent for telling the truth to have very best consequences; and so would not believe that it would have very best consequences unless he believed that the other would take the information as true rather than as false. Also being highly rational, the other would know this, and would not so take the information unless he believed that the informant believed he would so take the information. And this, of course, the informant would know. He could reason that if the other would take his information as true rather than false, it might have very best consequences to tell the truth, and that if he supposed that the other would so take his information and concluded that it would have very best consequences to tell the truth, then there would be good reason for the other so to take the information. But (as both would know), the informant could equally well reason that if he supposed that the other would take his information as false and concluded that it would have very best consequences not to tell the truth, then there would be good reason for the other to take the information as false.[9]

Before giving Hodgson's critics a hearing, I will make a preliminary assessment of Hodgson's argument. The argument may be represented by means of the following matrix (where "O_1"–"O_4" denote the four possible outcomes).

	B takes what he is told as true	B takes what he is told as false
A tells the truth	O_1	O_2
A does not tell the truth	O_3	O_4

How are the values of the outcomes related to each other? What Hodgson explicitly claims in his argument is that:

(i) The value of O_1 is better than that of O_2. (*A's* telling the truth has better consequences if *B* takes what he is told as true than if he takes it as false.)

and that:

(ii) The value of O_4 is greater than that of O_2. (*B's* taking what he is told as false has better consequences if *A* does not tell the truth than if he tells the truth.)

[9] Hodgson, op. cit., p. 44. Both Hodgson and his critics assume that the informant always knows the truth. They also assume that it is always better (in utilitarian terms) to believe what is true than what is false. I will later question both these assumptions.

How are the values of the other outcomes supposed to relate to each other? (Hodgson does not explicitly tell us, and (i) and (ii) imply nothing concerning that.) But, evidently, Hodgson must assume that:

(iii) The value of O_1 is greater than that of O_3. (*B*'*s* taking what he is told as true has better consequences if *A* tells the truth than if he does not.)

For otherwise—if the value of O_3 were at least as great as that of O_1—*A*'*s* not telling the truth would be (analogous to) what in game theory is called a *weakly dominating* strategy: given (ii), it would have better consequences than its alternative if *B* took what he was told as false, and, given the negation of (iii), it would have at least as good consequences if, instead, he took what he was told as true.[10] Being highly rational, *A* would of course know this and would therefore—contrary to what Hodgson assumes—know that AU prescribed that he should always not tell the truth.

Likewise Hodgson must assume that:

(iv) The value of O_4 is greater than that of O_3. (If *A* does not tell the truth, the consequences of *B's* taking what he is told as false are better than those of his taking it as true.)

For otherwise—if the value of O_3 were at least as great as that of O_4—*B*'*s* taking what he is told as true would be a weakly dominating strategy: given (i), it would have better consequences than its alternative if *A* told the truth, and, given the negation of (iv), it would have at least as good consequences if, instead, he did not tell the truth. Being highly rational, *B* would of course know this and would therefore—contrary to what Hodgson assumes—know that AU prescribed that he should take what he is told as true. (Moreover, given the truth of both (i) and (ii), the negation of both (iii) and (iv) would mean that, according to AU, *A* should not tell *B* the truth, although *B* should take what he is told as true. It would be rather astonishing if this combination of strategies always had the best consequences.)

Now, from (i)–(iv) it follows that:

(v) The value of O_1 is greater than both that of O_2 and that of O_3, and the value of O_4 is greater than both that of O_2 and that of O_3.

And this means that both O_1 and O_4 are (utilitarian analogues to) what in game theory is called *equilibrium* outcomes: Given that *A* tells the truth, *B*'s taking what he is told as true has better consequences than his taking it as false, and, given that *B* takes what he is told as true, *A*'*s* telling the truth has better consequences than his not telling the truth. Likewise, given that *A* does not tell the truth, *B*'*s* taking what he is told as false has better consequences than his taking it as true, and, given that

[10] In game theory, an agent's strategy *S* is weakly dominating if, and only if, its outcome is (i) better for him than the outcome of any alternative to *S,* given at least one combined choice of strategies by the other agents and (ii) as good for him as the outcome of any alternative to *S* given all other combined choices. Applying the notion to AU, "better for him" is replaced by "better", period, and "as good for him" is replaced by "as good", period.

B takes what he is told as false, *A*'*s* not telling the truth has better consequences than his telling the truth.[11] This, in turn, means that if either O_1 or O_4 obtains, both *A* and *B* do what AU prescribes.

What about the relation between O_1 and O_4 then? Hodgson's thesis requires that

(vi) The value of O_1 is the same as that of O_4. (The consequences of *A*'*s* telling the truth and *B*'*s* taking what he is told as true are just as good as the consequences of *A*'*s* not telling the truth and *B*'*s* taking what he is told as false.)

This does not directly follow from (i)–(v), but (v) entails that either (a) exactly one of O_1 or O_4 is optimal, or (b) they are equally best. If, however, (a) were the case, *A* and *B* would not, contrary to Hodgson's claim, be confronted with a co-ordination problem. Being highly rational, they would know that (a) was the case and would act so as the secure the optimal outcome (whichever that was). The value of O_1 is, therefore, the same as that of O_4, and these outcomes are equally best (optimal).

This, at any rate, is what Hodgson claims. Is the claim acceptable? This depends crucially on how *A*'*s* second alternative, viz. his not telling the truth, is to be understood. Hodgson assumes that AU prescribes that one should (at least often) successfully communicate the truth, and he claims that in the AU-society there are two equally good ways of doing that: telling the truth and taking what one is told as true, and not telling the truth and taking what one is told as false. Now suppose that *B* asks *A* what time it is, and that *A*, knowing that it is three o'clock, tells *B* that it is two o'clock. Suppose further that *B* takes what he is told as false. Here we have a case of not telling the truth—in the sense of telling *something incompatible with the truth*—and taking what one is told as false.[12] But it is not successful; it results in *B*'*s* believing that it is not two o'clock, which is of course true, but not in his believing the more specific and interesting truth that it is three o'clock. This he would have believed if, instead, *A* had chosen the first way of communicating the truth, that is,

[11] In game theory an outcome is in equilibrium (is an equilibrium outcome) if, and only if, for each agent, the consequences of his unilateral defection from the chosen strategy is not better for him. When applying the notion to AU, "not better for him" is replaced by "not better", period. We might call this "equilibrium in the *weak* sense", and define a notion of equilibrium in the *strong* sense by replacing "not better (for him)" by "worse (for him)" in the above formulation. It should be noted that O_1 and O_4 are equilibrium outcomes even in the strong sense.

[12] It might seem that this example—as well as other similar examples given by Hodgson, myself, and other commentators, which will be introduced later on—begs the question. For according to AU, since truth-telling will not exist in the AU-society, no communication will take place there. It is, however, almost unavoidable not to use such examples when discussing Hodgson's claims. One way out would be to follow Peter Singer's advice and stipulate that just before the events assumed in the examples take place "everyone in an until-then-normal society is miraculously converted to act-utilitarianism" (Peter Singer, "Is Act-Utilitarianism Self-Defeating?", *Philosophical Review*, 81 (1972): 94–104; p. 97. It is assumed that the habit of talking with one another would linger on for a while. But this proposal gives too much ground to Hodgson's opponents: if habits from the pre-utilitarian society, such as talking with one another, still exercise their influence on people in the AU-society, so might the habits of taking what one is told as true or expecting that promises will be kept. A better expedient, it seems to me, is to take such examples as being preceded by a tacit counterfactual: "assuming that people in the AU-society were to communicate with each other, then …"

had told him the truth, and *B* had believed what he was told. Since in many cases it is important to learn the more specific truth, AU prescribes in these cases the first way of communicating the truth rather than the second way. If, therefore, not telling the truth is understood as just telling something incompatible with the truth, the value of O_4 is often less than that of O_1.

If Hodgson's claim that the value of O_4 is the same as that of O_1 be acceptable, not telling the truth must be understood as telling *the denial (*or *negation) of the truth*. If, for example, *A* instead tells *B* that it is not three o'clock, and *B* takes what *A* says as false, then *B* will believe the more specific truth that it is three o'clock. It seems plausible that the consequences of this way of communicating the truth that it is three o'clock are just as good as those of *A*'*s* telling that it is three o'clock and *B*'*s* believing what he is told. Given this understanding of not telling the truth, it seems *prima facie* reasonable that, quite generally, the value of O_4 is exactly the same as that of O_1. And this is how I will I will understand Hodgson's notion of not telling the truth. (In the following, I will refer to the two interpretations as the *contrariety* and the *contradictory* interpretations, respectively, of "not telling the truth".)

Both Hodgson and his critics seem to assume that people in the AU-society would have no or very little reason to deceive each other. Hodgson himself does not even envisage this possibility, and in his critique of Hodgson Peter Singer says that "if everyone were an act-utilitarian most of the reasons, selfish or unselfish, which we would otherwise have for lying would not exist".[13] But this is not true, as I will try to show in some detail. (Most of my examples apply to act-utilitarians both in and outside the AU-society. Since, however, lying is not considered immoral in the AU-society, utilitarians in this society have more often (utilitarian) reasons to lie than have utilitarians outside it.)

Act-utilitarians do not, of course, have *selfish* (justificatory) reasons for lying. But they have other, specifically *utilitarian* ones—given that the lies were (rightly estimated to be) believed.[14] There are, for example, many cases where telling the truth is more harmful than lying, since believing the truth makes people unnecessarily upset (angry, sorry, depressed, etc.). The classic example of (what might be called) "beneficial deceit" is that of the doctor who dishonestly tells his patient that his condition is hopeful. In other cases telling the truth is indirectly harmful because, by making the addressee upset, it causes him to act suboptimally. Thus telling the person attacked by a tiger that he has only one cartridge left in his rifle may make him so nervous that he misses his target.

A third kind of cases where AU justifies telling lies are cases where the person addressed holds certain erroneous beliefs which cannot be easily corrected. Suppose,

[13] Peter Singer, op. cit., p. 100. A non-angelic act-utilitarian can, of course, have selfish *motivating* reasons for acting, although he cannot have selfish *justificatory* reasons for doing what he does. I take it that what Singer means is that the act-utilitarian lacks selfish justificatory reasons, not that he lacks selfish motivating reasons. Although the latter interpretation would strengthen Singer's case, the former has the clear advantage of being true.

[14] Remember that the following examples are taken to be preceded by the tacit counterfactual mentioned in the last note but one.

for example, that *A* wants his employee *B* to finish some important work by lunchtime the next day. As *B* is about to leave the office today, *A* asks him if the work is finished. It is not, but *B* knows that there will be plenty of time for him to finish the work tomorrow without detriment to his other tasks in the office; he also knows that *A* will not believe this, and will therefore order *B* to work overtime today if *B* tells him the truth. Since *B* does not want to miss tonight's important football match between Djurgården and AIK, he lies to *A*, knowing that his lie will not be detected. (This is a case where the lie has a selfish motivation but, it may be assumed, a utilitarian justification.)

A fourth kind of cases where AU holds that it is justified to lie are those where you can save several people by sacrificing one person with the help of a lie, or, more generally, cases where you can bring about good consequences for some people by bringing about (a lesser amount of) bad consequences for other people. Consider the following version of a stock example of this kind. (The incident is supposed to take place in our AU-society.)

> You are standing on a bridge watching a trolley hurtling down the tracks below you toward five innocent people. The brakes have failed, and the only way you can stop the train is to impede its progress by throwing some heavy object in its path. You yourself are not heavy enough, but luckily there is a fat man standing on the bridge next to you, and you could easily push him over the railing and onto the tracks below if only he takes another step forward. If he does and you then push him, he will die, but the five men will be saved. The fat man wants to take a step forward in order to have a better view, but, being aware that you are a fellow utilitarian, asks before he takes the further step, "You will not push me over the railing, will you?" "Of course not", I say with a reassuring smile. The fat man—who, like Hodgson and his critics, thinks that people in the AU-society have no or very little reason to deceive each other—is completely reassured and takes the further step. I push him over the railing, thereby saving the five men on the tracks.

It might be objected that I do not need to lie in the above situation; since the fat man is supposed to be a utilitarian, I only need to explain the situation to him, and he will gladly volunteer. But, first, there may be no time for me to explain. And, secondly, talking to him will probably not be effective: not even a convinced utilitarian might be sufficiently motivated to sacrifice his life at a moment's notice, however much this is demanded by his morality.[15]

Because of the amount of justified lying in the AU society the third of the propositions discussed above is not true: sometimes O_3 is better than O_1, for, as we have seen, given that *B* takes what he is told as true, *A*'*s* not telling the truth sometimes has better consequences than his telling the truth. For the same reason proposition

[15] Saying this is not to take a stand on the vexed issue whether *akrasia* really exists, that is, whether anyone ever voluntarily acts contrary to what he thinks he ought to do. If *akrasia* does not exist, then, on some meta-ethical views, not being able to force oneself to do what one thinks that utilitarianism says one should do shows that one does not "really" think that one ought to do it. But not being able to force oneself to do everything that one thinks utilitarianism says one should do does not show that one is not "really" a utilitarian.

(iv) is not true: given that *A* does not tell the truth, *B*'*s* believing what he is told sometimes has better consequences than his not believing it.

Nor are propositions (i) and (ii) true: common knowledge of the fact that (iii) and (iv) are not generally true undermines their validity too. Suppose that *A* knows that *p* is the case, but that *B*'*s* believing that non-*p* is the case has better consequences than his believing the truth. Suppose also that *A* (correctly) believes that *B* believes that *A*, for utilitarian reasons, wants to deceive him by lying to him: *B* will therefore take what *A* says as false. *A* therefore tells *B* that *p*, whereupon *B* believes that *p* is false, that is, that non-*p* is the case. Hence (i) is not true. (By means of a similar example it can be shown that (ii) is not true either.) Further, if (i)–(iv) are not true, nor are (v) and (vi): sometimes O_1 and O_4 are not in equilibrium and are not both optimal.

Does the fact that in certain cases propositions (i)–(iv) are false undermine Hodgson's claim that people in the AU-society are confronted with an unsolvable co-ordination problem? No, it does not. For, firstly, the propositions are false only in cases where AU prescribes deceit. But, obviously, in most situations of communication AU does *not* prescribe deceit, so in most cases the propositions are true and hence give rise to a co-ordination problem. This means that communication of true information in the AU-society is blocked even when it is prescribed by AU.

Secondly, the truth of the propositions is a sufficient, not a necessary, condition for there being a co-ordination problem: the problem exists also in many cases where (i)–vi) are false, that is, in many cases of deceit. Let us divide such cases into two main groups. The first group consists of those cases where *B* does *not* believe that *A* wants to deceive him. Those cases obviously pose the same problem as do cases where *A* does not aim at deceiving *B: B* does not know whether what is stated is meant to be taken as true or false, and *A* does not know whether it is best to tell the truth or to lie.

The second group consists of those cases where *B believes* that *A* wants to deceive him. Those cases pose no co-ordination problem given that, in addition, (a) *A* believes that *B* believes that *A* wants to deceive him; (b) *A* believes that *B* does not believe that (a) is the case; and (c) *B* believes that *A* does not believe that *B* believes that *A* wants to deceive him. If (a)–(c) obtain and, say, *p* is true, the rational thing for *A* to do is to tell *B* that *p*. For, as *A* (correctly) believes, it is then rational for *B* to believe that *p* is false, that is, to believe that non-*p* is the case. In all other cases that belong to the second group there is a co-ordination problem.

In nearly all cases of communication of information, whether true or false, in the AU-society, there is then an unsolvable co-ordination problem. In nearly all cases, therefore, the rational thing for *A* to do is to keep quiet, and, if *A* cannot refrain from talking, the rational thing for *B* to do is to suspend judgment.

Truth-Telling in the AU Society: Objections

The critics of Hodgson's theses have concentrated on his argument concerning truth-telling in the AU-society, in some cases treating the one concerning promise-keeping merely by implication. Since different critics raise different objections to Hodgson's arguments I will discuss the objections separately, beginning with those attacking the argument concerning truth-telling.

(i). *Gauthier's Objection*

David Gauthier discriminates between the two interpretations of "not telling the truth" mentioned in the preceding section.[16] He notes that, according to the contrariety interpretation, O_1 is better than O_4 (the outcome of *A*'*s* telling the truth and *B*'*s* taking what he is told as true is better than that of *A*'*s* not telling the truth and *B*'*s* taking what he is told as false); given this interpretation, there is, therefore, no problem of co-ordination. But, Gauthier concedes, according to the contradictory interpretation, O_1 and O_4 are both optimal, and in this case there is a co-ordination problem. But, he thinks, this problem can be solved by means of the notion of *salience*.

> Stating that-p and believing what is stated is [a] more direct way of communicating that-p than stating that-non-p and believing the negation of what is stated. Hence salience attaches to the outcome of telling the truth and believing what is told. A second argument will reinforce this conclusion. There are circumstances in which it is possible to verify whether stating that-p is to be taken as a way of communicating that-p, or a way of communicating that-non-p. I say "The cat food is in the cupboard and the cat is not in the kitchen," and you look and see whether the cat food is in the cupboard and the cat not in the kitchen, or whether the cat food is not in the cupboard and the cat is in the kitchen. If act-consequentialists tell the truth in these situations, they thereby make telling the truth and believing what is told salient, not just for such situations but in general. In this way they develop the practice of communicating information by telling the truth.[17]

According to Gauthier, then, stating the truth and believing it is salient, since it is "a more direct way" of communicating the truth than stating the opposite of the truth and believing its negation. What then does it mean that the former is "more direct" than the latter? (Gauthier does not tell us.) Presumably that it is simpler and therefore requires less mental effort. But why is this thought to make it more *salient* than the latter? Presumably because it is held that people, *ceteris paribus*, prefer what is simpler and requires less mental effort to what is less simple and requires more mental effort. But this is certainly not a universal truth: people often have the opposite preferences. Would they have it in the present case? Hodgson thinks that

[16] David Gauthier, "Coordination", *Dialogue: Canadian Philosophical Review*, 14 (1975): 195–221; repr. in David Gauthier, *Moral Dealing: Contract, Ethics, and Reason* (Ithaca and London: Cornell University Press, 1990): 375–97; my references are to the reprint.

[17] Ibid., p. 293. (By "act-consequentialists" Gauthier understands both act-utilitarians and adherents of ethical egoism.)

they would have preferences of both kinds and that these would cancel each other out. In a passage seemingly anticipating the present objection he says:

> The *difficulty* of telling a lie in a non-act-utilitarian society arises mainly because of the need to tell a 'good' lie, in order to avoid both detection and the bad consequences of someone's being misled. In our postulated society, unless the act-utilitarian principle required the truth to be told, there would be no need to prevent detection, and no question of anyone's being misled; and so the lie would not have to be a 'good' one. A minimal degree of inventiveness might perhaps still be required to tell a lie; but we may assume that our rational act-utilitarians would have this, and that if any disvalue were involved in the effort required to use this inventiveness, it would be balanced by the satisfaction of exercising the skill.[18]

Who is right, Hodgson or Gauthier? Before addressing this question I want to point out that if Gauthier is right, O_1 and O_4 (see above) are not, contrary to what he assumes, both optimal: the preferences for the former outcome tip the balance and make O_1 the unique best outcome. Rational act-utilitarians would therefore opt for O_1, not because it is salient, but because it is the unique best outcome. But would they?

It should be admitted that it is not easy to decide whether Hodgson or Gauthier is right on this issue: our knowledge of the mental make-up of fully rational act-utilitarians is far from complete. There are, however, certain considerations that tell in favour of Hodgson's position. (These considerations, it should be noted, are relevant to both of Gauthier's arguments set forth in the above quotation from him.)

In a society such as ours, where there is usually good reasons to believe what other people tell us, and people have formed habits to believe accordingly, it often requires more effort not to believe what someone says than to believe it. But there is no reason to think that people form these habits in our AU-society. (Simply to assume that people in this society habitually believe what other people tell them begs the question.) And what else could make it the case that it required more effort from people in this society not to believe what other people told them than to believe it? (It might be replied that some things that people tell us seem intrinsically more plausible than their contradictories and are therefore more easily believed than disbelieved. This is true, of course, but then other things that people tell us seem intrinsically *less* plausible than their contradictories.)

But all this is actually beside the point. What Gauthier says suggests that he thinks that the alternatives are (i) believing what someone says and (ii) believing its contradictory. But obviously there is a third alternative, viz. that of suspending belief. And if there is (almost) just as much reason to believe what someone says as to believe its contradictory, this alternative seems to be the unique rational epistemic stance to take. Since our act-utilitarian agents *ex hypothesi* are highly rational, this

[18] Hodgson, op. cit., p. 43. Cf. what Rawls calls "the Aristotelian Principle": "other thing being equal, human beings enjoy the exercise of their realized capacities (their innate and trained abilities), and this enjoyment increases the more the capacity is realized, or the greater its complexity." (John Rawls, *A Theory of Justice*, p. 426.)

is then the stance they are expected to take—and this is so even if it would require somewhat more effort than just believing what they are told.[19]

The crucial question is whether people in the AU society would interpret the situations they confront according to the contrariety or the contradictory interpretation?[20] If both *A* and *B* interpret a situation in the former way (and know of each other that they interpret it that way, and so on), then *A* will tell *B* the truth and *B* will take what he is told as true. (I temporarily disregard the kinds of situations, mentioned in the previous section, where *A*, for utilitarian reasons, lies to *B*.) Evidently there are situations which they would interpret according to the contrariety interpretation. To be a situation of that kind, what is required is that *A* believes that *B* believes that there are more than two alternatives any one of which might be true (and that *B* believes that *A* believes that, and so on). If in such a situation *A* tells *B*, concerning one of these alternatives, that it is true, then, as *A* knows, *B* has a reason to think he is told the truth. The alternative way of communicating the truth consists in *A*'*s* telling *B* what is not true and *B*'*s* taking what he is told as false. But, as both can easily verify, if the (according to *B*) possibly true alternatives are more than two, the former way of communicating the truth is more likely to be successful. And, as they both know, this gives *A* a reason to choose it.

If, for example, as *B* believes, *A* believes that *B* believes that either *C*, *D*, or *E* has committed the crime, and *A* knows who did, the rational way of communicating the truth to *B* is telling the truth. If, however, *B* only suspected two people, the situation must be interpreted according to the contradictory interpretation, and both ways of communicating the truth would be equally good.

The fact that people in the AU-society would sometimes view the situations they confront according to the contrariety interpretation certainly weakens Hodgson's position, but not very much. For there are many situations that would be interpreted according to the contrary interpretation. Moreover, the kinds of situations where *A* has utilitarian reasons for lying to *B* are rather frequent. Hence, the situations where *B* really has good reasons to believe that *A* tells him the truth would probably not be many.[21]

---

[19] This is also Donald Regan's opinion. "But the easiest thing of all is to avoid the question of how to take *A*'s remark, by ignoring it entirely." (Regan, op. cit., p. 35.)

[20] As I pointed out in Sec. 2 above, Hodgson evidently interprets the situations according to the contradictory interpretation.

[21] C. Provis, op. cit., objects to Gauthier's solution to the following co-ordination problem: *A* and B want to meet each other either at *x* or at *y*, no matter at which place. For some reason, going to *x* is the salient option. Gauthier suggests that each agent should restrict his possible actions to (i) seeking the salient outcome and (ii) ignoring it. The agent seeks the salient outcome by going to *x* and ignores it by randomizing on an equal basis between going to *x* and going to *y*. It is easily seen that both agents seeking the salient outcome, that is, going to *x*, has a higher expected value than both agents ignoring salience. Being rational, they therefore both go to *x*, thus successfully co-ordinating their actions.

Provis objects that Gauthier's depiction of the situation is inadequate: the agents have a third alternative, viz. seeking the *non-salient* outcome, that is, going to *y*. And if both go to *y*, they likewise successfully co-ordinate their actions and will meet each other.

(ii). *Singer's Objections*

Peter Singer raises three objections to Hodgson's argument.[22] In the first, Singer asks us to consider the case of an office clerk *B* living in the AU-society, who on a particular day intends to work overtime. *B's* only means of transportation home is by bus, and for some reason it is very important that he does not miss the last bus. *B* asks his colleague *A* when the last bus departs. *A* knows the answer, and he also knows that it would have best consequences to inform *B* of it. But what answer should he give? If Hodgson is right, Singer says, it is not more likely that *A* can inform *B* by telling the truth than by telling a lie. But, Singer objects, Hodgson is wrong: there is a utilitarian reason for *A* to tell *B* the truth. If *A* tells a lie, then, whether or not *B* believes it, he will (in all probability) not go to the bus stop in time. But there is a fifty-fifty chance that *B* will take the answer given by *A* to be true and, therefore, if told the truth, will go to the bus stop in time.[23] This gives *A* a reason for telling *B* the truth. And, Singer says, this reason is actually a fairly strong one:

> Once there is some reason for A to tell the truth, there is more than enough reason for him to do so. For B, being highly rational, will have thought of the considerations just pointed to, and will be aware that there is a reason for A to tell him the truth, and A will know this, and so on. So we get the Hodgson spiral working in the other direction, and A will have the normal utilitarian reason for telling the truth—that is, that B will take the information to be true and make arrangements based on its truth[24]

This objection presupposes for its validity the contrariety interpretation of Hodgson's argument, viz. that *A* has to choose between telling the truth or telling any of a number of false answers. Given the contradictory interpretation—which, as I have argued above, is the one that Hodgson must subscribe to—the objection obviously does not work. If, as Singer says, "there is a fifty-fifty chance that *B* will take the answer given by *A* to be true", there is a fifty-fifty chance that he will take it as false. If, therefore, the truth is that *p*, and *A* says that not *p*, there is a fifty-fifty chance that *B* will take the answer as false and hence believe that *p*, that is, believe the truth.

As a matter of fact, Singer's objection might be mistaken even given the contrariety interpretation. What is wrong with both Singer's and Mackie's criticism of

---

Provis is, no doubt, right. But, as far as I can see, Gauthier could accept Provis's objection and still have a good case—if, that is, he can defend his view concerning the importance of salience against my objections to it.

[22] Peter Singer, op. cit.

[23] It might be objected that the same beneficial consequences will also obtain, with the same probability, in certain cases of telling *B* a lie, viz. in those cases where *A* tells *B* that the bus will depart some time *earlier* than it actually will. For, if in those cases *B* believes that *A* tells him the truth, he will *not* miss the bus. In defence of Singer we might stipulate that coming to the bus "in time" means coming to the bus stop either just when the bus is due to depart or "shortly before" (admittedly a vague expression). If *B* arrives earlier than "in time", then (we might further stipulate) the consequences are worse than if he arrives "in time": he prefers go on working in his office to waiting for the bus longer than a "short while" before it is due to depart.

[24] Singer, op. cit., p. 98.

Hodgson, Donald Regan says, is the assumption that the behaviour of an agent which is supposed to constitute a piece of communication

> will be taken by the other agent who perceives it to have *some* communicative effect. But there is no reason to assume that there will be any communicative effect at all.[- - -] Singer completely ignores the possibility that *B* will pay no attention whatever to what *A* says.[25]

(This accords with what I said above concerning Gauthier's objection.)

Singer's first objection to Hodgson's argument concerned the special case where information is given as an answer to a question. But in most cases information is given without its having been requested. In his second objection Singer turns to such cases, that is, to cases where information is volunteered. He takes as an example a situation where a stranger *A* comes up to "me" and says: "There is a very good film on the local cinema this week." In such a case, Singer claims, I have a good reason to believe that *A* tells me (at least what he thinks is) the truth.

> Since by going through the business of inventing what *A* says to me—thinking to myself, "He says the film is good, but he may be telling a lie, so the film may be bad"—I am no more likely to arrive at the truth than if I take what *A* says at face value, why should I bother to invert it? Am I not just a fraction more likely to take it at face value? If I am, *A*, being highly rational, will know this, and will know that he is more likely to produce best consequences if he tells the truth, while I, being highly rational, will know this, and so expect *A* to tell the truth. […] and so we get the spiral unspiraling again, and we have all the reason we need for telling the truth.[26]

The crucial step in this argument is the claim that it requires more mental effort to disbelieve what someone says than to believe it ("why should I *bother* to invert it?"—my italics). The cogency of this claim was discussed in the preceding subsection and will not be repeated here.

According to Singer's third objection, if there were no social practices of truth-telling and promise-keeping in the AU-society, there would be act-utilitarian reasons for taking steps to establish such practices.

> Any steps toward the formation of these practices would have the good consequences of making desirable activities possible. Since telling the truth and keeping promises could help in the formation of these practices, while lying and breaking promises could not, this would give an additional reason for telling the truth and keeping promises. The spiraling effect would come into operation. This would ensure the rapid development of the practices. The informer or promisor would then have the dual reasons of preserving the useful practice and fulfilling expectations.[27]

As Singer notes, Hodgson is aware of the objection and tries to meet it; there are, Hodgson says, act-utilitarian reasons against taking the steps that would establish such practices. But, as Singer also notes, it is far from clear what these reasons are taken to be. The following is what Hodgson says by way of argument:

---

[25] Donald Regan, op. cit., p. 35. (As for Mackie's criticism of Hodgson, see the next subsec.)

[26] Singer, op. cit., p. 100.

[27] Ibid., op. cit., p. 101.

> Such steps could have good consequences, but, although perhaps justified by act-utilitarianism, they would amount to a partial rejection of act-utilitarianism and so would be inconsistent with our assumptions. These steps would amount to a partial rejection of act-utilitarianism, because the persons would be forming habits to do acts known not to be justified according to act-utilitarianism; and they would form these habits only if they resolved to refrain from applying act-utilitarianism in relation to these acts.[28]

Singer suggests that, according to the most plausible interpretation of this passage, what is claimed to be contrary to AU are the *initial* steps, those taken before there is any established practice and, therefore, any expectations. (I do not think that this is what Hodgson has in mind—I will come back to that later on—but that does not really matter; the important thing is whether, thus interpreted, the passage can be used to meet Singer's third objection, not whether it is what Hodgson had in mind.) This interpretation, Singer thinks, is supported by the fact that in a later chapter of his book Hodgson raises a similar objection to the act-utilitarian justification of the practice of punishing: although an unbroken record of punishment may deter potential offenders, there is no act-utilitarian reason for starting the record in any given case rather than in the next.

Singer objects that this reasoning is built on the faulty assumption that the only consequences of an action are those for which the action is a necessary or a sufficient condition. But, Singer says, an action may have consequences—"may contribute to a result"—for which the action is neither a necessary nor a sufficient condition, and such consequences are, of course, relevant to AU.[29]

> The contribution that my vote makes toward the result I judge to be best in an election is a relevant consideration in deciding whether to vote, although it is, almost certainly, neither a necessary nor a sufficient condition of that result; for if this were not so, the act-utilitarian view would leave us with a result which was unconnected with the actions of any of the voters, since what is true of my vote is equally true of any individual vote. [- - -] In the cases we were considering originally, an act of telling the truth or keeping a promise will normally have greater utility than would its opposite, because it has a reasonable chance of contributing to the beneficial consequences of setting up a desirable practice.[30]

According to Singer, then, if in an election my vote for a candidate is considered as not making a contribution to the candidate's being elected, this outcome would be "unconnected with the actions of any of the voters, since what is true of my vote is equally true of every individual vote".[31] Evidently Singer finds the conclusion of this reasoning paradoxical and takes it to show that, therefore, my vote must be considered as making a contribution to the outcome. But I cannot find anything paradoxical in the claim that in cases of overdetermination, like the one we are now

---

[28] Hodgson, op. cit., p. 48.

[29] Singer, op. cit., p. 103.

[30] Ibid.

[31] Singer says that my voting for the successful candidate is, "*almost* certainly, neither a necessary nor a sufficient condition of that result" (my italics). Since the version of AU that Hodgson discusses is primarily concerned with *actual* consequences (see above p. 1), "almost" should be omitted: to be relevant, the case we are dealing with must be a case of *actual* over-determination, not one of *possible* over-determination.

considering, the outcome is causally unconnected with *any* individual action—although, of course not, with *all* of them. That is why, if in the above voting situation I had to choose between voting and performing another action, AU would prescribe that I ought to perform the other action even if my candidate's being elected was very valuable, whereas the consequences of the other action was only of little (positive) value.

In the case of participating in starting a practice of promise-keeping or truth-telling there is not, as in the case of casting one's vote, a certain definite outcome that one either is, or is not, instrumental in achieving; rather it is a matter of participating in the gradual realization of something—let us call it "trust"—among members of one's society. (So the parallel drawn by Singer is quite misleading and of little help for his argument.) But the contribution that a single individual can give to establishing the practices of promise-keeping or truth-telling in his society is probably negligible: the practice would most certainly be established (and preserved) whether or not, say, individual *A* contributed. Notice that in our present society some people are untrustworthy, and known to be so, without this tending to abolish the practices of promise-keeping and truth-telling. (It might be objected that *A's* not keeping his promises or telling lies is usually directly harmful to those who trust him. But this objection overlooks the fact that, if Hodgson is right, nobody trusts *A* in the AU-society. If, on the other hand, and contrary to what Hodgson argues, the AU-society would be sufficiently transformed into a non-AU society for other people to trust him, then there are utilitarian reasons for *A* to match up to the trust. See below section "Truth-telling and promise-keeping in non-AU-societies".)

But, someone might protest, even if there might be a practice of promise-keeping or truth-telling in a society although *A* does not participate in it, it would be minimally better, in utilitarian terms, if he did. Similarly, even if *A's* contribution to starting such a practice might not be necessary, it would still be better if he participated; for then the useful practice would be established (if only minimally) faster. So there is a utilitarian reason for each person to contribute to establishing the beneficial practices of promise-keeping and truth-telling.

Hodgson's answer, I think, would be the following (and this is how I think the above quotation from him should be interpreted): In our AU-society, establishing the practices in question means taking steps to form habits of keeping promises and telling the truth, habits having roughly the same scope and strength as those prevailing among members of actual societies. The taking of such steps by at least the great majority of people in the society would certainly have good consequences—it would establish the practices in question, and these are very useful—and it would "perhaps", that is, if AU concerned itself with *collective* actions, be justified by AU. But the habits in question are habits to do *individual* actions, which are wrong according to AU.[32] So if any person in the AU-society acts so as to form the habits,

[32] In Sec. 5.5 above, I pointed out and discussed some problems connected with taking C as a (wholly or partly) collective moral theory.

he acts contrary to what AU prescribes. As Hodgson says: "they would form these habits only if they resolved to refrain from applying act-utilitarianism to these acts."

(iii). *Mackie's Objection*

J. L. Mackie suggests that members of the AU-society could manage without using sentences having truth-values; instead they could use what he calls "belief-imperatives", imperatives of the form "Believe that p". (They could even simply use "p" instead of the longer form.)[33] Mackie makes two claims on behalf of this device. The first is that there is no great psychological difficulty involved in believing what one is thus ordered to believe: "For most things that it is in accordance with utility for people to believe are truths".[34] The second claim is that this device is better than the ordinary practice of truth-telling.

> This use and acceptance of belief-sentences may not be exactly what we ordinarily call the communicating of information, but it is practically equivalent to this, and in some ways superior to what we have in all actual societies. For though we have conventional rules of truth-telling, we well know that they are often violated, occasionally for benevolent reasons but more often in support of divergent interests. Would it not be better to be sure that your neighbour was always telling you to believe what it would be best for you to believe than to be uncertain whether he was telling you the truth or deceiving you for his private ends and against your interests?[35]

I disagree with both of these claims. As I argued at length in section "Truth-telling in the AU-society: Hodgson's argument" above, I think that there are many kinds of situation where AU says that you should lie to other people. Recognizing this, people in the AU-society would often hesitate to believe what they are ordered to believe—even if they believed that believing it and acting upon the belief would have better consequences than not believing it. In the AU-society there would, for example, be many situations where AU prescribed that the interests of one person should be sacrificed for the sake of better consequences on the whole. In some such situations one would have to deceive somebody in order to accomplish this. Of course, no miscalculation being suspected, the victim would think that the sacrifice was justified. But, unless human nature were drastically transformed, people in the AU-society would live in constant fear of suddenly being sacrificed *ad maiorem gloriam utilitatis*. In not a few cases, then, there would be "great psychological difficulty involved in believing what one is thus ordered to believe". So Mackie's first claim is false.

As for the second claim, the claim that the device proposed is better than the ordinary practice of truth-telling, it is evident that the proposal does not even meet Hodgson's objection. Hodgson claims that when *A* states that *p*, *B* does not know

---

[33] J. L. Mackie, "The Disutility of Act-Utilitarianism", *Philosophical Quarterly*, 23 (1973): 289–300.

[34] Ibid., p. 297.

[35] Ibid., p. 298.

whether *A* wants to communicate that *p* is the case or that non-*p* is the case. It seems that *B* has exactly the same problem if *A*, instead, would say, "Believe that *p*". *B* knows that *A* either asks him to believe that *p* is the case or that non-*p* is the case, but one is not more probable than the other. Switching from the indicative to the imperative mood does not solve the problem posed by Hodgson.[36]

(iv). *Lewis's Objection*

David Lewis illustrates Hodgson's thesis by means of the following example. Two highly rational act-utilitarians, "you" and "I", are put in separate rooms, each having a red and a green button at his disposal. If, and only if, we both push either the red or the green button, we bring about the Good; otherwise we bring about the Bad. We know all this, we know that we know, and so on. You manage to send me a message, "I pushed red". Do I then have a reason to push red?[37]

Not if Hodgson is right. According to Hodgson, Lewis says, I must reason as follows.

> I have not the slightest reason to believe you unless I have reason to believe that you think that I have reason to believe you. But I know that you—knowledgeable and rational creature that you are—will not think that I have reason to believe you unless I really do. Do I? *I cannot show that I have reason to believe you without first assuming what is to be shown: that I have reason to believe you.* So I cannot, without committing the fallacy of *petitio principii*, show that I have reason to believe you. Therefore I do not. Your message gives me not the slightest reason to believe that you pushed red, and not the slightest reason to push red myself.[38]

But this is absurd, Lewis says, so there must be a flaw in the argument. Lewis thinks that the flaw comes with the step taken in the italicized sentence, the step where, he says, I tacitly assumed that my reason to believe you must be found only in facts about the situation and us, "our utilitarianism and rationality, our knowledge of these, our knowledge of one another's knowledge of these, and so on".[39] But, Lewis asks, why must my reason to believe you be limited to these facts? To show that I have such a reason, I could start with any premise that gives me a reason to believe you, provided it is available to me and consistent with, as well as indepen-

---

[36] Donald Regan's criticism of Mackie's proposal is more radical. "If verbal communication were not established, people would not make vocal noises with the intent to communicate, and the only vocal noises I would hear would in fact flow from other motives." (Regan, op. cit., p. 36.) This accords with what I said concerning Gauthier's objection in Subsec. 3: (i) above.

[37] David Lewis, "Utilitarianism and Truthfulness", *Australasian Journal of Philosophy*, 50 (1972): 17–19; p. 17. The example seems to presuppose the contrariety interpretation, since each of us has four alternatives: push red, push green, push both buttons, and push neither button. But, since we are utilitarians and know that pushing either button has greater expected utility than pushing either both buttons or neither button, we take only the former alternatives into consideration. So the example actually presupposes the contrary interpretation.

[38] Ibid., p. 17. A similar argument, Lewis points out, could be applied to promising: "for an example of this, just change the message in my example to 'I will push red'" (ibid.).

[39] Ibid., p. 18.

dent of, the facts about the situation and us. And, Lewis says, there is actually such a premise.

> The premise that you will be truthful (whenever it is best to instill in me true beliefs about matters you have knowledge of, as in this case) is just such a premise. It *is* available to me. At least, common sense suggests that it would be; and our only reason to suppose that it would not is the Hodgsonian argument we are disputing. […] On the one hand it is *consisten*t with our rationality and utilitarianism, our knowledge thereof, and so on. […] On the other hand, it is not *implied* by our rationality and utilitarianism, our knowledge thereof, and so on.[40]

Admittedly, the premise *seems* available to me—at least if "you think it is best" is substituted for "it is best". As Lewis says, common sense suggests that it is. But, first, common sense has little experience of thinking as a highly rational act-utilitarian: even if the premise is available to common sense, it might be unavailable to the utilitarian. And, secondly, common sense might easily be misled by what is omitted in Lewis's presentation of the example: Lewis never mentions that you and I are members of the AU-society; it is therefore easily imagined that we are two act-utilitarians living in a predominantly non-utilitarian society, who usually deal with, and are accustomed to, people who adhere to the norms concerning truth-telling prescribed by CSM.

As far as I can see, however, Lewis's objection begs the question at issue. Of course, if I accept the premise, then I must, on pain of contradiction, believe what you said. (For I know that (you know that) it would be best to instill true beliefs in me.) But what reason do I have to accept the premise?[41] If Hodgson is right, the premise is false, and I have no reason to accept it. Lewis does not prove that Hodgson is wrong, he just takes it for granted.[42]

(v). *Hoerster's Objection*

Yet another critic of Hodgson is Norbert Hoerster.[43] Hoerster argues that it is possible, as well as justified on act-utilitarian grounds, to introduce the practice of truth-telling in the AU-society allegedly lacking it. At first one gets the impression that Hoerster wants to defend the *collective* version of act utilitarianism (see section

---

[40] Ibid., p. 18.

[41] Lewis says that I *know* that the premise is true. (The reason to believe your message and, therefore, to push red myself is, he says, "premised on further knowledge that I do in fact possess (ibid., p. 19).) True, if I *know* that the premise is true, then, as a matter of conceptual truth, I must have good reasons for believing it and should be able to state them, but Lewis does not tell us what these reasons are.

[42] Adrian Piper (op. cit.) criticizes Lewis's attempt (as well as a similar attempt by Allan Gibbard in his (unpublished) Ph.D. dissertation) to refute Hodgson. If I have understood Piper correctly, his main criticism of Lewis (and Gibbard) is that they beg the question by assuming what Hodgson implicitly denies, viz. that communication of any sort would be possible in the AU society. I think that Lewis could meet the objection. It is true, he might retort, that my example seemingly presupposes that people in the AU society are able to communicate with each other. But so do the examples given by Hodgson himself. In neither case, however, is such an assumption really made. The examples should be read: "assuming that people in the AU society *were* able to communicate with each other, then ..."

[43] Norbert Hoerster, "Is Act-Utilitarian Truth-Telling Self-Defeating?", *Mind*, 82 (1973): 413–16.

"Truth-telling in the AU society: objections": (ii) above). For he says that a member *A* of the AU society is *pro tanto* obliged to tell the truth if, by telling the truth, *A* contributes to creating the expectations necessary for establishing the practice of truth-telling in his society. And this condition, Hoerster says, is ordinarily satisfied.

> For if A *and all other individuals in similar situations* will actually tell the truth, whenever, but for the consideration of turning truth-telling into a practice, its utility is indifferent, then truth-telling will become more frequent than lying and as a result people will generally expect to get true rather than false statements from their fellow beings.[44]

But later Hoerster seems to recognize that this answer does not address the objection raised by Hodgson, which is directed against individualistic act utilitarianism. For he imagines that Hodgson might object to the above answer by pointing out that "[e]ach separate extra true statement will only 'contribute' to creating a new expectation, if it does not remain alone".[45] So Hoerster turns his attention to this objection, that is, to an objection against AU actually entailed by what Hodgson says.

It is not easy, however, to say what exactly the answer that Hoerster gives to Hodgson is. Hoerster admits that a single act of truth-telling cannot be justified on act-utilitarian grounds by its alleged contribution to establishing the practice of truth-telling in the society. For whether this practice will ever be established is, as he admits, just the point at issue.

> The solution lies, however, in describing the specific act to be tested by the act-utilitarian formula not as telling the truth on some occasion, but rather as creating some individual's expectation to be told the truth in the future. Thus each individual can be shown to have an obligation usually to tell the truth to those of his fellows whom he frequently contacts. And, in this way, the practice of truth-telling is established, at least in its most important aspect, namely between people not being strangers to each other.[46]

I take it that what Hoerster wants to claim is the following: Suppose that *A* and *B* are two members of the AU-society who regularly meet each other and also have some need to communicate with each other. At the beginning, they do not trust each other to speak the truth. But, by starting to be truthful to *B*, *A* gradually causes *B* to trust him to speak the truth. Each single act of *A's* telling *B* the truth contributes to this happy result. There is therefore an act-utilitarian reason for *A* to be truthful to *B* on almost every occasion. And what is true of *A* is true of all or most members of the AU-society in relation to their nears and dears.

I disagree. There are several reasons why *B* will not come to trust *A*.[47] One is that in many cases *B* does not have the opportunity to find out, without much ado, whether *A* speaks the truth or not. In order to do that, *B* must make some extra effort. And why should he do that if he initially does not trust *A* and therefore does not care

---

[44] Ibid., p. 414; my italics.

[45] Ibid., p. 415.

[46] Ibid., p. 416.

[47] In the rest of this section I will, for the sake of brevity, speak of "trusting someone", instead of using the longer expression "trusting someone to speak the truth".

about what *A* says? Secondly, even if *A* speaks (what he thinks is) the truth, and *B* tries to check out whether what *A* has said is true, *B* may still believe that *A* has told what is false. For either *A* or *B*, or both, may be mistaken concerning relevant facts: *A* may falsely think he speaks the truth, and/or *B* may falsely think that what he is told is false. And, thirdly, even if *B* is quite sure that *A* has spoken the truth, this, as he is aware of, does not give him a reason to trust *A* in the future. For, as Hodgson points out, there is no reason to think that members of the AU-society do not sometimes tell the truth: they might sometimes tell what is false, sometimes what is true.

A fourth reason why *B* will not trust *A* is that *A* sometimes, on strict utilitarian grounds, will lie to *B*. As we saw in section "Truth-telling in the AU-society: Hodgson's argument" above, there are several kinds of situations where AU prescribes that people tell lies—on condition that those whom they are addressing trust them. So if *A* thinks that *B* trusts him, he will lie to *B* if he thinks that they find themselves in such a situation. If *B*, whether truly or falsely, does not believe that they find themselves in such a situation, he will believe what he is told. If later he finds out that *A* lied to him, his trust in *A* will be diminished. Moreover, it may be questioned whether, as Hoerster thinks, *A* really has "an obligation usually to tell the truth to those of his fellows whom he frequently contacts". There are several reasons why *A* does not have such an obligation. First, if Hodgson is right, *A* has at most an obligation (a utilitarian reason) to *communicate* the truth to other people. Now, as been repeatedly mentioned, there are two ways of doing this: telling the truth and not telling the truth. And if Hodgson is right, one way is as good as the other.

Suppose, however, that Hodgson is wrong: it is better to communicate the truth by means of telling it than by means of not telling it. But this does not mean that *A usually* has an obligation to tell *B* the truth. For, secondly, there are, as I have repeatedly said, many cases where AU prescribes that *A* should lie to *B*. Moreover, as I argued above, *B* will not trust *A*. So even in cases where it would be better, in utilitarian terms, that *B* believed what is true than what is false, it might be impossible for *A* to accomplish this. Given that "ought" implies "can", *A* does not then have an obligation (a reason) to communicate (or to tell) *B* the truth. At most, therefore, what can be said is that *A* has an obligation sometimes to *try* to communicate the truth to *B*. But, unfortunately, we know that he will most probably fail.

And even if people in the AU-society would succeed in establishing the practice of truth-telling between people who regularly meet each other and need to communicate, all is not well. The AU-society would still be severely handicapped in comparison with our society. For even if it is true, as Hoerster claims, that it is most important that there is a practice of truth-telling among non-strangers, it is also very important that the practice extends to strangers too. In all but the most primitive kind of society people need to trust even those whom they are not acquainted with. And Hoerster's proposal, as he himself admits, cannot establish that this is the case in the AU-society.

Promise-Keeping in the AU Society: Hodgson's Argument

What is true of truth-telling is, Hodgson claims, also true of promise-keeping. In the AU society promising would be pointless, so there would exist no such practice.[48] For the only reason for keeping a promise in this society is that it would have best consequences to do what one has promised. But if it would not have best consequences to do the action apart from the promise, it would not have best consequences given the promise. The fact that one has promised to do an action would add to the utility of performing the action only if this fact gave rise to additional expectations that the action would be done. But if the promisee knows that the promisor is an act-utilitarian, knows that the promisor knows that the promisee knows that, and so on, then the fact that the promisor has promised to do the action does not increase the promisee's expectations that the promisor will do it. And, of course, the promisor knows that, knows that the promisee knows that the promisor knows, and so on. In Hodgson's own words:

> So, a promised act could have greater (comparative) utility (than it would have had if it had not been promised) only if the promisee has a greater expectation that it would be done (than he would have had if it had not been promised); but there would be a good reason for such greater expectation only if (in the promisor's belief) the act would have such greater utility. Being highly rational, the promisor would know that the greater expectation was a condition precedent for the greater utility; and so would not believe that the act would have greater utility unless he believed that the promisee had greater expectation. Also being highly rational, the promisee would know this, and so would not have greater expectation unless he believed that the promisor believed that he had greater expectation. And this, of course, the promisor would know.[49]

The argument that Hodgson gives concerning promise-keeping in the AU-society has then a different structure than that concerning truth-telling. The argument is roughly that in the AU-society the consequences of keeping a promise are not better than that of not keeping it. So there is no point in promising in the AU-society, and there will therefore be no such practice. But is this really the whole story? Does not promising involve saying something that is either true or false, or does it not at least give rise to a (true or false) belief? If so, it seems that the Hodgsonian argument concerning truth-telling in the AU-society applies to the practice of promising as well. I will try to show that this is really the case.

Suppose that *A* and *B* are members of a non-AU-society, and *A* tells *B*, "I will visit you tomorrow". Let us assume that both take this as a promise. Does it have a truth-value? No, says the received view: a promise is a performative (a performative utterance), and performatives are neither true nor false.[50] I think that the received view is mistaken in the case of promising, but I will not argue this here. It suffices

---

[48] Let us in the following by the practice of promising understand the practice of making, as well as (normally) keeping, promises.

[49] Hodgson, op. cit., p. 41.

[50] See, e.g., Justus Hartnack, "Performative Utterances", in Paul Edwards (ed.), *The Encyclopedia of Philosophy*.

for my purpose if it be admitted that, if *B* trusts *A*, what *A* says to *B* gives rise to a belief in *B* to the effect that *A* will visit him tomorrow, and that *A* knows that it does. *B*'*s* belief is either true or false: if *A* visits *B* tomorrow, the belief is true, otherwise it is false. Whether or not *A* visits *B*, *B* will for some time have entertained exactly the same belief that he would have entertained if instead another person whom he trusts, had said, "B will visit you tomorrow", which is clearly not a performative.[51]

If, then, promising normally gives rise to beliefs, the ways of communicating true beliefs by promising in the AU-society is essentially similar to that of ordinary communication: either (i) the promisor makes a promise that he will keep, and the promisee believes that he will keep it, or (ii) the promisor makes a promise that he will not keep, and the promisor believes that he will not keep it. For reasons similar to those adduced in section "Truth-telling in the AU-society: Hodgson's argument" above, the promisee would have no way of knowing whether the promisor will or will not keep the promise. So there would be no point in promising in the AU-society, and the practice of promising would not exist.

By two different routes, then, we arrive at the same conclusion. As far as I know, there have been four replies to this conclusion, three of which attempt to refute it, and one which concedes it but thinks that it is no objection to AU. I shall discuss these replies in turn, starting with the latter.

But before that I shall introduce a helpful division of different kinds of promises, a division borrowed from Russell Hardin.[52] Hardin divides promises into three categories: (i) *exchange promises*, promises that "facilitate exchanges made over time"; (I give you now my *x* and you promise to give me your *y* later.) (ii) *co-ordination promises*, promises that "facilitate our getting together or otherwise accomplishing some joint venture"; (iii) "*gratuitous" promises*, promises that "are the promissory equivalent of acts of beneficence: there is no quid pro quo, no evident benefit to the promisor".[53]

Promise-Keeping in the AU Society: Objections

(i). *Mackie's Objection*

Mackie concedes that promising would have a very restricted role in the AU-society sketched by Hodgson; its only role there, Mackie claims, is to help people make combined choices between (actual or presumed) utility maxima—that is, to make choices between several outcomes of combinations of actions, each

[51] In the case of promising, the performative justifies a belief that "the performer" will perform another action—e.g., paying a visit to the promisee—such that its non-performance shows that something was wrong with the performative. In the case of most other performatives—such as apologizing, naming, and inviting—something similar is not true. This, I think, explains why, e.g., promising has a communicative effect different from, e.g., that of apologizing.

[52] Russell Hardin, *Morality Within the Limits of Reason* (Chicago and London: The University of Chicago Press, 1988).

[53] Ibid. p. 60.

combination consisting of one action by each of several agents, outcomes which are, or are considered to be, equally best.[54] (I will come back to this role in the next subsection.) But, Mackie says, Hodgson is wrong in inferring that a society, which (almost) completely lacked the practice of promising, would therefore necessarily be in dire straits. The other purposes served by promising in our society would be served by other means in the AU-society: it would be served by the utilitarian morality of its inhabitants.

> The main point of promising in our present society is that it enables people with divergent aims to co-operate to some extent. The point of a promise is that it helps to construct a compromise. Alf, say, would like best to get his work done and not pay any wages for it; Bill would like best to be paid wages and do no work; but Alf would rather pay wages and get the work done than not get the work done and pay no wages, and Bill would rather work and be paid than neither work nor be paid. They can reach a compromise between their divergent interests if Alf promises to pay Bill if he first does the work, and Bill trusts this promise, and does the work, and Alf then keeps his promise and pays Bill. But if there had been no divergence between their aims, if they had each been concerned only for their common welfare, no promising or promise-keeping would have been necessary.[55]

I have two comments to make on this. First, Mackie's example does not show that act-utilitarians can do equally well without promising. True, if "they had each been concerned only for their own common welfare", Bill could perhaps trust Alf to pay him when he has finished his work. But in the AU-society, although there is "no difference between their aims", Alf and Bill are certainly not "concerned only for *their own common* welfare"; being utilitarians, they are concerned for *the general* welfare. So, if Alf, before he pays Bill, would realize that he could spend his money in some way that had better consequences than handing it over to Bill, he would do that. Being an act-utilitarian. Bill would not, of course, have any reason to complain. But knowing that Alf is an act-utilitarian, he would not trust Alf to pay him for his work. And, if it were important for him to get money for work done, he would probably never start working for Alf unless being paid in advance—but it may be doubted that Alf would trust Bill sufficiently to pay him in advance. (Bill's decision not to work for Alf is consistent with AU. It might, of course, be the case that it would have better consequences if Bill worked for Alf and Alf gave the money to someone else than if he gave them to Bill. But it might also be the case that it would have even better consequences if Bill worked for someone else, whether or not Alf hired someone else to do the job or did not get the job done.)

My second point concerns Mackie's claim that "[t]he main point of promising in our present society is that it enables people with divergent aims to co-operate to some extent". Of course, some promises are of this semi-contractual nature, for example, the one in Mackie's example. But there are other kinds of promises, for example, gratuitous promises (see the end of section "Promise-keeping in the AU society: Hodgson's argument" above. Whereas co-ordination promises are often, at least partially, made (and kept) out of (mutual) self-interest (the promisor wants to

[54] J. L. Mackie, op. cit.

[55] Ibid., p. 296.

secure some benefit for himself), the latter are made (and kept) out of "other-interest" (the promisor wants to secure some benefit for someone else). Examples of such promises are promises to assist one's friend with money in case of emergency, to visit a sick relative before long, to discuss this appendix with me when it is finished, and so on. Due to the discrepancy between the promisee's welfare and the general welfare, such promises could seldom be trusted in the AU-society. They would therefore be pointless and, hence, not made—to the detriment of the members of the society.

Now if the promising practice is useful but, as Mackie concedes, has a very restricted role in the AU society, it may be wondered whether there could not be some replacement for it. (Act-utilitarians, it should be noted, would not object to transforming a kind of situations in such a way that the outcomes of the actions required by AU in the transformed situations were better than the outcomes required in the untransformed situations, the cost of transformation included.) Such a replacement would guarantee, for example, that Alf pays Bill for his work even if in the untransformed situation it is better that Bill works and Alf does not pay him than that Bill works and Alf pays him. Now whereas in some cases co-ordination promising may be replaced by other (more clumsy) devices—Alf, say, paying Bill *continuously* during the time the latter works—the chances of finding out such replacements for gratuitous promising seem bleak. This is due to, what may be called, the "non-reciprocity" of such promises: in the case of these promises, as distinguished from co-ordination promises, it is not the case that the promisor should keep his promise on condition that the promisee does something else. The promisee has therefore no hold on the promisor like the one Bill has on Alf, no device, that is, which could be used as a means for creating a replacement for such promising. In the case of gratuitous promising there is no essential feature of the promising situation that can be used to restructure the value-ordering in such a way that it becomes better, in utilitarian terms, to keep the promise than not to keep it.

(ii). *Narveson's Objection*

As we saw, Mackie thinks that the only role that promising would play in the AU society is to help people to make *combined* choices between several (actual or presumed) utility maxima. Jan Narveson thinks that, in addition, promising would help people to make *individual* choices between such maxima.[56] And these two roles, he claims, are far from being restrictive; together they range over the whole field of promising as ordinarily practiced. Moreover, he says, there would be just as much promising in the imagined AU society as in actual societies.[57] Let us consider the two alleged roles for promising, starting with the second.

---

[56] Jan Narveson, "Promising, Expecting, and Utility", *Canadian Journal of Philosophy*, 1 (1971): 207–23.

[57] One gets the impression that Narveson thinks that promises of these two kinds are rather frequent. As far as my experience goes, they are rather uncommon, especially promises of the latter kind.

Sometimes, Narveson says, an agent finds himself in a situation such that he believes that two or more alternatives have equally best outcomes. Suppose that *B* believes that a utilitarian *A* is in a situation where actions a_1–a_n appear to have equally best outcomes. Then *B* cannot possibly know which of these actions *A* will do; barring further information, he knows only that, for each action, there is a possibility of 1/*n* that *A* will do it. Suppose further that it is important for *B* to know what *A* will do in order to be able to know what is best for himself to do. There is, then, a utilitarian reason for *A* to pick one of these actions, promise to do it and do it.

We get the same result in the case of combined choices, Narveson says. Suppose that two people want to have lunch together. There are *n* possible restaurants to meet at. Each person wants very much to have lunch with the other, but neither cares which restaurant they go to. Suppose further that the best outcome (in utilitarian terms) is that they meet at some—no matter which—of the *n* restaurants. There is, then, a utilitarian reason for them to pick one of the restaurants, promise each other to go there and keep the promise. Picking one of the restaurants and agreeing to go there is, of course, using an arbitrational device, but, Narveson asks rhetorically,

> are we to infer from the fact that some arbitrational device is necessary here that such a device requires an extra-utilitarian justification, or that it has the status of a moral rule, independent of its utility? I should think not—any more than adherence to the rules of cricket or baseball needs a justification independent of *their* utility.[58]

There is, Narveson says, no other role for promising, whether in the imagined AU-society or in actual societies. If there is no uncertainty concerning what an agent will do apart from the promise—if, for example, we (correctly) believe that a utilitarian agent has one best action with a best outcome at his disposal (and he knows that we believe this, and we know that he knows, and so on)—there is no use for promising, and it will not take place.

> If everybody in a group knows exactly what everyone else, and himself, wants, at all times, then the supposition that an institution of promise-keeping, with its attendant language of obligation, would be of any use at all would be quite baseless.[59]

I think that Narveson is mistaken in several respects. First, he arbitrarily restricts the range of promising to situations where there are several utility maxima. Thus he overlooks both the kind of semi-contractual promises that Mackie thought are the most important and gratuitous promises.[60] For an example of the former kind of promises, see Mackie's example cited above. For an example of the latter, suppose that you have to pay the rent for your flat within a week or be evicted, but you do not have the money. Having learned of your financial trouble I promise to give you the needed money. This example, as well as the one given by Mackie, is two of countless promises that do not satisfy the several-utility-maxima condition.

---

[58] Ibid., p. 225.

[59] Ibid., p. 227.

[60] The promises considered by Narveson may be seen as solutions to co-ordination games situations, whereas those considered by Mackie may be seen as solutions to co-operative games situations, especially PD-situations. (Gratuitous promises fall outside the scope of game theory.)

Would any of the kinds of promises not recognized by Narveson take place in the AU-society? Semi-contractual promises would not, there being no difference between people's aims in the AU-society. Moreover, in the case of many such promises, the outcome agreed upon is not necessarily optimal (in utilitarian terms), taking into account only the welfare of the agents in question instead of the general welfare. (The forming of economic cartels does not usually benefit the public at large.) As for the existence of gratuitous promising in the AU-society, I can only repeat what I said in the preceding subsection: due to the discrepancy between the promisee's welfare and the general welfare, such promises could seldom be trusted in the AU-society. They would be pointless and, hence, not made.

There are, then, many cases of promising in actual societies that would not take place in the AU-society. There are also many cases of promises that are normally kept in actual societies, but would not be kept in the AU-society. This is something that Narveson cannot admit, since in his opinion what has given rise to, and justifies, the norms concerning promise-keeping is our (unconscious?) utilitarian convictions. The only merit of Hodgson's argument, he says, is that, by making us reflect on the matter, "it does help us to see more clearly the role of promising in human affairs, and likewise to see that its obligational force is entirely a function of its utility".[61] This is, I think, a second issue about which Narveson is mistaken.

To see this, consider a case where something unforeseen happens between the giving of the promise and the time when it is to be kept. If the unforeseen thing had not happened, then, all things considered, the outcome of keeping the promise would have been better than that of not keeping it. Since it happened, however, the two outcomes are equally good, and both keeping and not keeping the promise are therefore equally right according to AU. I do not think, however, that this is what most people think, especially not if what raised the value of the outcome of not keeping the promise were benefits to the promisor. According to CSM, a promise should not, of course, be kept come what may; but it should be kept unless the consequences of keeping it are considerably worse than those of not keeping it, especially if keeping it is important to the promisee.[62]

My third objection to Narveson is that he begs the question at issue. Consider the two kinds of promises which are most favourable to Narveson's position, and the only kinds of promises whose existence he recognizes, viz. those made in situations where several of the agent's (agents') possible actions appear to have equally best outcomes. In such situations, whether they occur in actual societies or in the AU-society, there is, Narveson states, a utilitarian reason to pick one of the actions, promise to do it, and do it. But according to Hodgson there is no such reason in the AU-society. Hodgson's argument for his claim, let us recall, is (roughly) the following: As both *A* and *B* know, there would be a reason for *A* (the promisor) to do *a* (the

[61] Ibid., p. 227.

[62] This is also the standard view of deontologists. Primarily, of course, they claim that it is the *true* view, but in their attempts to justify this verdict, they explicitly or implicitly claim that the true view is also the view of common sense. See, e.g., David Ross, *The Right and the Good*, *passim*, and *Foundations of Ethics*, pp. 87–113.

promised action) only if he believed that *B* (the promisee) would expect that. But, as *A* knows, *B* would expect that only if he believed that there was a reason for *A* to do *a*. Thus there is no utilitarian reason to do what one has promised to do just because one has promised to do it. Hence, members of the AU-society would not make any promises.

This argument, if valid with respect to any promises, is certainly valid with respect to (putative) promises to do one of several actions with equally best outcomes. In order to show that Hodgson is wrong, therefore, it is not enough just to asseverate that he is; one should show what is wrong with his argument.

(iii). *Gauthier's Objection*

David Gauthier does what Narveson fails to do: he tries to show what is wrong with Hodgson's argument.[63] According to Gauthier, "the primary function of the practice of promising is to serve as a device for coordination".[64] (Thus he takes the practice of promising to be equally restricted as Narveson does.) Suppose, Gauthier says, that you and I want to meet tomorrow. I promise you that I will be at my office at 2 p.m. Prior to the promising there were many best alternatives as to when and where to meet, but, acting independently of each other, we were unable to secure any of them. My promise helps us to single out one of the outcomes—to our mutual benefit.

> If I promise you that I will perform some action, then I make the outcome of my doing that action salient. In promising I change, not the utilities in the situation in which we find ourselves, but our conception of that situation. My reason for performing the act promised is that in the situation conceived in terms of the promise, the promised act leads to the unique best equilibrium outcome. And this reason is provided by the making of the promise, for in the situation apart from the promise, no act leads to a unique best equilibrium outcome.[65]

I think Gauthier is right to some extent: if *A* and *B*, two inhabitants of the AU-society, believe that they confront a co-ordination situation with more than two alternatives (and also believe of each other that they believe that, and so on), *A*'*s* promising to do his part for bringing about a certain one of the alternatives provides him with some reason to do it and *B* with some reason to expect it. (This is analogous to what was said with respect to truth-telling at the end of section "Truth-telling in the AU society: objections": (i) above.) This weakens Hodgson's position, but not very much. For, first, *A* knows that *B* knows (and so on) that *A* is a utilitarian and therefore holds that the only reason to keep a promise is that it maximizes utility. So even if, at the time of promising, *A* thinks that keeping the promise has better consequences than not keeping it, much that might happen until it is time to keep it could tip the balance. So *B*'*s* reason to trust *A* cannot be very strong. Knowing this, *A*'*s* reason for keeping the promise is not very strong either.

---

[63] David Gauthier, op. cit. Gauthier wants to show that "at least a rudimentary form of promising is not only possible but also rational and desirable for all act-consequentialists" (p. 294).

[64] Ibid., p. 295.

[65] Ibid. My promise, Gauthier says, does not change the utilities. This is, no doubt, true. But it changes the *expected* utilities, and that is what is important.

Moreover, if, as I argued in section "Promise-keeping in the AU society: Hodgson's argument" above, promising is a way of (knowingly) raising true or false beliefs in the promisee, promising inherits all the utilitarian reasons for deceiving people by lying to them discussed in section "Truth-telling in the AU-society: Hodgson's argument" above. (In the case of promising, deceit will, of course, be effected, not by (strictly speaking) lying, but by not keeping the promise.) Finally, and quite decisively, as I said in the previous subsection, promising to bring about a certain one of several best outcomes far from exhausts the field of promising; it is in fact only a small part of it.

(iv). *Gibbard's Objection*

Allan Gibbard bases his objection to Hodgson on David Lewis's theory of convention.[66] A crucial element in this theory is the claim that people will rationally keep an agreement if, and only if, it is common knowledge in their society, or just among the parties to the agreement, that people have generally kept their agreements in the past; in other words, people will rationally keep their agreements if, and only if, past history has established a *convention*. With one exception (to be introduced later in this section) this is so, Gibbard argues, even in the AU-society.

I do not have space to critically discuss at length Gibbard's complex and very subtle defence of his thesis. I shall content myself with pointing out some weak points in Gibbard's argument, points which, I think, cast grave doubts on the tenability of his position. But before that I want to point out that, even if tenable, Gibbard's objection does not cut very deep. So even if his objection were well-founded, it would not destroy, only to some extent weaken, Hodgson's thesis concerning the fate of promises in the AU-society.

Firstly, Gibbard too discusses only one kind of promises, viz. *co-ordination* promises (see section "Promise-keeping in the AU society: Hodgson's argument" above). And, as I said above, even if such promises were given and held in the AU-society, it does not follow that other kinds of promises too are given and held there. Secondly, as Gibbard himself admits, a condition for there being in the AU-society an agreement justified by AU is that "the parties share their relevant experience, […] [that] they agree on all probabilities relevant to the effects of making the agreement binding and carrying it out."[67] If this condition is violated,

> then it is as if they worked from different utility scales altogether. Even if they agree on their ultimate ends, they may disagree on what more immediate ends would foster those ultimate ends. […] In those circumstances, different people may find different agreements optimal.[68]

Since the parties do not always have the opportunity to make sure that they agree on all relevant probabilities, the condition is probably far from always fulfilled.

---

[66] Allan Gibbard, op. cit.; David Lewis, *Convention: A Philosophical Study* (Cambridge, Mass.: Harvard University Press, 1969).

[67] Gibbard, op. cit., p. 107.

[68] Ibid., p. 109.

So, as I said above, Gibbard's criticism at most weakens Hodgson's thesis. But does it even accomplish that? I do not think so. For there are, in my opinion, (at least) two unconvincing points in Gibbard's argument. One of these may be introduced as follows: As I said above, Lewis claims that it is rational to keep one's agreements if, and only if, past history has established a convention. But then, Gibbard points out, it may be objected that

> [i]f it is rational to follow a proto-convention only when it has a supporting history, then it must have been irrational for anyone to follow it in the first place, and so in a society of rational agents, a supporting history could never arise.[69]

But the objection is dismissed by Gibbard:

> I agree that if it is rational to follow the proto-convention only when it has a supporting history, then a supporting history could never arise in the first place. It may, however, have been rational to follow the proto-convention before it had a supporting history precisely because a history of its being followed could later make it rational to follow it and a history of its not being followed—an 'undermining history', I shall say—would later make it irrational to follow it.[70]

So it is rational to start a proto-convention because this is to start giving it a supporting history, which, in turn, makes it later rational to follow the proto-convention. And, Gibbard adds, to start following the proto-convention is rational because this has good expected consequences, whereas not to start following it has bad expected consequences.

> In the first situation to which the proto-convention applies, where it has neither a supporting nor an undermining history, parties' knowledge [of the situation] gives them reason for following the proto-convention, because establishing the beginnings of a supporting history has good expected consequences and establishing the beginnings of an undermining history has bad expected consequences.[71]

Now suppose that John and Harriet for the first time make an agreement to meet in the park at noon. (This is the example that Gibbard uses throughout his article.) To do what they have agreed to do has good expected consequences, Gibbard claims, since it contributes to establishing a supporting history. But, it might be objected, the contribution of a single case to this goal is, if at all noticeable, quite negligible. A supporting history will, or will not, become established whether or not, say, John keeps *this* agreement. (The same is true of every single agreement, whether or not a supporting history has been established at the time when the agreement is made. Consider how conventions survive occasional deviations from them in our actual non-AU-society.) This means that many circumstances that may attract John's attention after the agreement has been made will, for utilitarian reasons, tell against going to the park. Thus, for example, John might discover that one of his favourite

---

[69] Ibid., p. 101. A *proto-convention* is a convention without its supporting history. A *supporting history* for a convention is the common knowledge that the convention has been followed in the past.

[70] Ibid., p.101.

[71] Ibid., p. 102.

programs will be on TV at noon, or that it unexpectedly starts to rain just when he is about to go to the park. Harriet knows that such things might happen, and John knows that Harriet knows this, and so on. So if a supporting history will ever be established—which I doubt—the proto-convention it supports will be a very weak one and hardly to be much trusted. It will not be as useful as a corresponding convention in our non-AU-society.

But, it might be objected, I have left out of account a circumstance that definitely speaks in favour of John's going to the park, viz. that Harriet will go to the park as agreed and will be disappointed if John does not come. This circumstance, it might be hold, gives John a strong utilitarian reason (and a corresponding motivation) for keeping the agreement. But does it? I do not think so. Harriet is in exactly the same situation as John. So she too might become aware, after the agreement has been made, of circumstances that tell against her going to the park. And John knows this. Harriet knows that John knows, and so on. This further reduces both Harriet's and John's utilitarian reasons for going to the park: it is far from certain that the other person will come to the park. (This too is common knowledge between John and Harriet, and it further reduces their reasons for going to the park.) If the probability that the other will stay home is estimated to be higher than 0.5, the utilitarian rational thing for each of them to do is to stay home. And, as far as I can see, this might very well be the case.

I have hitherto presupposed that the best outcome is that of both John and Harriet going to the park, the next best that of both staying home, and the worst outcomes are those of only one going to the park. This is also how Gibbard describes the example:

> The best outcome they can achieve is to meet as agreed, and the next best outcome is for both to stay home and read. Because each would find it distressing to come and not find the other, the worst outcome they can achieve is for one to come to the park and the other to stay at home.[72]

Does this mean that John and Harriet going to the park would be better than both staying home even without an agreement to go to the park, or that it is better only given an agreement? I think that the first alternative, although it is in some respects the more plausible interpretation, is not the intended one: since it is common knowledge among John and Harriet that they are rational act-utilitarians who know about their situation, they do not have to make an agreement to do what both know is the best thing to do.

The second alternative is also problematic. Note that it presupposes that, without an agreement, both going to the park and both staying home are equally good. For, as we just saw, the first outcome cannot be better than the second. Nor can the second outcome be better than the first: if staying home is better than going to the park, it would be irrational to agree to go to the park. But if, without an agreement, going to the park and staying home are equally good, then it seems that John and Harriet should stay home. For meeting each other in the park at noon requires making an

[72] Ibid., p. 91.

agreement, but staying home is the default position and requires nothing of that sort. Since making the agreement and carrying it out involves unnecessary costs with respect to the time and effort spent, the rational thing to do is to abstain from making any agreement—which means staying home.

I have now dealt with what, in my opinion, is one of two unconvincing points in Gibbard's argument. The other point concerns Gibbard's claim that the conventions which, according to him, exists in the AU-society are not "conventional moral rules" in Hodgson's sense. If they were, Gibbard admits, his argument would not refute Hodgson:

> Hodgson stipulates that in the society we are to consider, there are no "conventional moral rules." If this rules out AU-conventions, then in showing that members of an openly act-utilitarian society would keep promises if they had an appropriate AU-convention, I shall not be refuting Hodgson.[73]

But, Gibbard argues, Hodgson's stipulation does not rule out AU-conventions:

> Something is a conventional moral rule in Hodgson's sense only if deviations from it "are generally regarded as lapses or faults open to criticism." AU-conventions do not need to be sustained by criticism. They are sustained by common knowledge that each person chooses the most favourable prospect when he acts, and that he reasons inductively. Hence there can be an openly act-utilitarian society which satisfies Hodgson's stipulation that there be no conventional moral rules, and which still has AU-conventions.[74]

AU-conventions, Gibbard then claims, differ from conventional moral rules in that the former do not have to be sustained by criticism of deviations from them, while the latter have to be thus sustained. This, I think, is wrong. Conventional moral rules are not primarily sustained by criticism of deviations from them—nor does Hodgson say so. What sustains them are successful moral indoctrination and/or moral insight. And the same is certainly true of AU-conventions. Moreover, deviations from conventional moral rules are, no doubt, criticized, but so are in all probability deviations from AU-conventions too. Thus, unlike Gibbard, I cannot see that there are any relevant differences between the AU-conventions and conventional moral rules.[75]

Truth-Telling and Promise-Keeping in Non-AU-Societies

As mentioned at the beginning of the appendix, Hodgson claims that also the correct application of AU by members of a predominantly non-AU society like our own would (probably) have worse consequences than would acceptance of CSM. Hodgson argues as follows: Let *A* be a highly rational act-utilitarian living in a

[73] Ibid., p. 98. An *AU-convention* is, roughly, a proto-convention such that it is common knowledge among those who are parties to it in the AU-society that it has been followed by them in the past.

[74] Ibid., p. 98.

[75] Hodgson's treatment of exchange promises (see Sec. 4 above) is criticized by Russell Hardin, op. cit., p. 61 f. But Hardin's discussion contributes nothing of interest.

non-AU-society. If those with whom he interacts know that *A* always tries, possibly successfully, to act in accordance with AU, many of the problems with respect to promise-keeping and truth-telling that affect members of the AU-society appear; even if other people tolerate *A's* moral conviction, they do not trust him, and, as we have seen, lack of trust causes great disutility.[76]

If, on the other hand, *A* does not avow his acceptance of AU, other problems appear on account of *A's* then having to deceive other people. He will feign to accept CSM and will conform to it in all cases where there is a risk that non-conformance will be detected. But then *A* is insincere: he only *seems* to recognize the notions of personal obligation, apologizing, and blaming, notions which are essential to CSM. All this deceit will destroy his candour and openness of character. And Hodgson concludes:

> There are alternatives which are open to those who accept personal rules approximating to the conventional rules of their society, but which are not open to those whose only personal rule is that of the act-utilitarian principle. Even though the former persons might not always choose, from the acts open to them, those with the best consequences, nevertheless the consequences of the acts which they do choose might be better than the best consequences which the latter persons could bring about through the alternatives open to them. Not only is this possible, but the cases we have considered suggest that it is probable.[77]

I disagree. Certainly, if *A* makes it known that he is an act-utilitarian, he is in trouble. But what if he does not? Of course, his candour will be destroyed. But this should not worry an adherent of AU. I fully agree with what Mackie says on this issue:

---

[76] In "Consequences of Utilitarianism", *Dialogue: Canadian Philosophical Review*, 7 (1969): 639–42, L. W. Sumner criticizes Hodgson's account of the fate of an act-utilitarian in a non-utilitarian society with respect to promise-keeping. It is possible for the promisee, Sumner contends, to shape the situation in such a way that a promise given by the act-utilitarian agent can be trusted. (Sumner assumes that it is common knowledge that the agent in question is an act-utilitarian.) All that the promisee has to do, Sumner says, is to put himself to some trouble having some disutility unless the promise is kept. "By doing so he may manufacture the conditions necessary for the act utilitarian to keep the promise." (p. 641) (Sumner evidently assumes that what the promisee does will be known to the promisor.)

This is, no doubt, a workable strategy. But, obviously, it does not make the act-utilitarian agent trustworthy, in the sense of "trustworthy" here under discussion, viz. being such as to keep his promises even when it does not maximize expected utility. Moreover, the suggested proposal does not square with Sumner's view that the disutility of the trouble that the promisee puts himself to because of his expectation that the promise will be kept "will ordinarily precede rather than follow the breaking (or keeping) of the promise and thus cannot be a consequence of it" (ibid.). (But the fact that Sumner's proposal does not square with this view does not really matter, since Sumner's view is not true: the *trouble* may precede the breaking of the promise, but the *disutility* is certainly a consequence of it.)

[77] Hodgson, op. cit., p. 58 f. Howard Sobel, who seems to agree with Hodgson, has pointed out an interesting analogy between Hodgson's argument and David Gauthier's argument for the rationality of constrained (egoistic) maximization. (Jordan Howard Sobel, "Kent Bach on Good Arguments", *Canadian Journal of Philosophy*, 19 (1989): 447–54.) As Sobel notes (p. 451), Gauthier states that "[t]he essential point in our argument is that one's disposition to choose affects the situations in which one may expect to find oneself". (David Gauthier, *Morals by Agreement* (Oxford: Clarendon Press, 1986), p. 183.

> A thorough-going act-utilitarian would be impervious to the social pressures that condemn even benevolent deceit; knowing that he is deceiving his fellows only for the sake of common good, he will feel a glow of conscious virtue each time he takes them in.[78]

So AU prescribes that *A* should not avow his adherence to AU but should conform to CSM whenever non-conformance might be detected.[79] And if he does, he will produce at least as much utility as he would have done had he accepted CSM instead. Therefore AU is not self-defeating for its adherents in actual, non-AU-societies.

Suppose, however, for the sake of argument, that Hodgson is right: it would have been better, on act-utilitarian grounds, if *A* had accepted CSM instead. This means that AU is, in Derek Parfit's terms, *indirectly individually self-defeating* in *A's* case. A moral theory *T* is thus self-defeating, Parfit says,

> when it is true that, if someone tries to achieve his T-given aims, these aims will be, on the whole, worse achieved.[80]

It is evident that if, according to AU, *A* ought to have accepted CSM instead of AU, then AU is often self-defeating in this sense for its adherents in non-AU-societies. However, as Parfit argues, being indirectly individually self-defeating does not tell against a moral theory. An agent's less than optimal achievement of his *T*-given aims is not the result of his doing what *T* tells him to do. It is the result of his being disposed to act in a certain way, often due to his belief in *T*. And if this is the case, these dispositions and this belief are not sanctioned by *T*. If *T* tells him anything in this respect, it tells him to change his dispositions and belief, and thereby his aims, if he can, and adopt other ones. Thus AU would tell *A* to adopt the motivations and beliefs which, of those open to him, would have the best consequences.[81] It may be the case, however, that the agent cannot change his dispositions and moral beliefs. Perhaps, for example, *A* cannot stop believing in AU. This is, however, just a sad fact about reality, comparable to *A's* being, say, poor. For if A had been rich instead of poor, he would, let us assume, have produced more value than he does now. In neither case is AU to be blamed.

So whether or not an act-utilitarian living in a predominantly CSM-society abides by his moral theory or abandons it in favour of CSM, it creates no problem for AU.

But, someone may ask, why is AU *directly* (collectively) self-defeating, and not only *indirectly* (individually) self-defeating—and thus innocuously self-defeating—for people in the AU-society as well. In that case too, it might be claimed, the less than optimal outcomes of people's actions are the result of their having beliefs and dispositions not sanctioned by AU. The answer is that in this case the less than optimal outcome of an action is not, as in the case of act-utilitarians living in non-AU-societies, the result of the *agent's* beliefs and dispositions; it is the result of

---

[78] Mackie, op. cit., p. 299.

[79] For a classic defence of this position, see Henry Sidgwick, *The Methods of Ethics,* pp. 485–92.

[80] Parfit, op. cit., p. 5. *Directly* individually self-defeatingness is introduced at the end of Sec. 3.4.

[81] Cf. Parfit, op. cit., Sec. 18.

other people's beliefs and dispositions, notably their beliefs concerning the agent's beliefs and dispositions. To verify this, suppose that *A* and *B* are two inhabitants of the AU-society. One day *A* rejects AU and becomes an adherent of CSM. If, however, *B* believes that *A* is still an adherent of AU (and *A* believes that *B* believes this, and so on), what AU prescribes that *A* should do with respect to telling the truth and keeping his promises to *B* is obviously no different from what it would have been if *A* had still been an adherent of AU. And there is no reason to think that the converted *A* produces more utility than the unconverted *A* would have done. So *A's* conversion was not prescribed by AU.

Concluding Remarks

In this appendix I have discussed D. H. Hodgson's claims that AU is self-defeating, both for people in the AU-society and for act-utilitarians in non-AU-societies. I have defended the first claim but argued against the second. When defending the first claim, I have also briefly analysed and tried to assess the arguments adduced by Hodgson, viz. that truth-telling and promise-keeping would be absent in the AU-society. In addition, I have pointed out the great amount of deceit, prescribed by AU, which would take place in the AU-society, something that strengthens Hodgson's position.

For the most part, I have been busy defending Hodgson's first claim. The defence has almost entirely consisted in criticizing the objections that have been raised against it. I am unpersuaded by these objections, and have tried to show, for each of them, what is wrong with it. Of course, even if, in each case, my criticism is well-founded, this does not show that Hodgson is right; there might be other objections, waiting to be raised, that would show that Hodgson's claim is false. But, although my defence of the claim does not amount to a proof of it, I hope that my defence has strengthened its plausibility.

But, it may be asked, assuming that Hodgson's claim is true, what does this show? Not very much, it seems. Why could not an adherent of AU with equanimity accept that his morality is self-defeating for people in the AU-society, a society that will never be realized? The important thing, he could say, is that Hodgson's second claim is not true: it is not true that AU is self-defeating for people in actually existing societies.

I disagree. If Hodgson's first claim is true, this shows that AU is not, as its adherents claim, the *fundamental* moral theory. The fundamental moral theory is not, I take it, self-defeating in any (logically) possible world. But this claim is controversial and has to be argued for at some length. I shall therefore content myself with a less controversial claim that serves my purpose equally well: the fundamental moral theory is not self-defeating in that possible world where everyone accepts the theory and nearly always does what the theory tells them to do. In this case it cannot plausibly be held that the theory's being self-defeating shows that something is wrong with the world rather than with the theory—unless knowledge that other people

hold the same moral convictions as oneself is taken as showing that something is wrong with the world.

The truth of Hodgson's first claim, together with the falsity of the second claim, also indicates that AU is *parasitic* on CSM, in the sense that the former is "successful" with respect to truth-telling and promise-keeping only if most people conform to the latter.[82] For, as we have seen, an act-utilitarian *A* living in a non-AU-society ought, according to AU, to conform to the norms of CSM with respect to truth-telling and promise-keeping. And the consequences of *A's* conforming to CSM in these respects are, on the whole, beneficial. If, on the other hand, *A* had lived in the AU-society, where no one adheres or conforms to CSM, he would have been unable to bring about such beneficial consequences.

It might be objected that this inability does not depend on people's general *adherence* to AU, but on their common *knowledge* of this adherence; if everyone in the AU-society thought that everyone else adhered to CSM, then everyone, though still adhering to AU, would nearly always conform to the norms of CSM, thus bringing about the beneficial consequences of truth-telling and promise-keeping. The short answer to this objection is that even in this case is AU parasitic on CSM: the beneficial consequences of anyone's conforming to AU with respect to truth-telling and promise-keeping depend on everyone's conforming to CSM.

The discussion in this Appendix has been concerned with two moral practices, truth-telling and promise-keeping. In both cases, I have argued, AU is self-defeating unless most people conform to CSM. The question is whether this result can be generalised. In other words, is act-utilitarianism *generally* parasitic on common-sense morality? I discussed this question in Chap. 7.

[82] It is not, of course, required that they conform to CSM *in toto*, only that they conform to its norms concerning truth-telling and promise-keeping.

Bibliography

Acton, H.B., ed. 1969. *The Philosophy of Punishment*. London: Macmillan.

Albee, Ernest. 1902. *A History of English Utilitarianism*. London: Swan Sonnenschein.

Annas, Julia. 2004. Being Virtuous and Doing the Right Thing. *Proceedings and Addresses of the American Philosophical Association* 78: 61–75.

Anscombe, G.E.M. 1958. Modern Moral Philosophy. *Philosophy* 33: 1–19.

———. 1990. Modern Moral Philosophy. *Ethics* 101: 42–63.

Aristotle. *Nicomachean Ethics* (Find references).

Ashford, Elisabeth, and Tim Mulgan. 2008, Fall. Contractualism. In *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. http://plato.stanford.edu/archives/fall2008/entries/contractualism/.

Augustine. 1948. The Lord's Sermon on the Mount. In *Ancient Christian Writers: The Words of the Fathers in Translation*, ed. Johannes Quasten et al. Westminster: Newman Press.

Baier, Kurt. 1958. *The Moral Point of View*. Ithaca/New York: Cornell University Press.

Baron, Marcia. 1997. Kantian Ethics. In *Three Methods of Ethics: A Debate*, ed. Marcia Baron, Philip Pettit, and Michael Slote, 3–91. Oxford: Blackwell.

Baron, Marcia, Philip Pettit, and Michael Slote. 1997. *Three Methods of Ethics: A Debate*. Oxford: Blackwell.

Beauchamp, Tom. 1998. Editor's Introduction. In *An Enquiry concerning the Principles of Morals*, ed. David Hume's. Oxford: Oxford University Press.

Beck, Lewis White, (ed. and tr.). 1976. *Immanuel Kant: Critique of Practical Reason and Other Writings in Moral Philosophy*. Reprint. New York: Garland.

Bennett, Jonathan. 1995. *The Act Itself*. Oxford: Oxford University Press.

Bentham, Jeremy. 1823. *Not Paul but Jesus*. Issued under the pseudonym "Gamaliel Smith, Esq". London: John Hunt.

Bentham, Jeremy. *An Introduction to The Principles of Morals and Legislation* (Find references).

Bergson, Henri. 1977. *The Two Sources of Morality and Religion*. Trans R.A. Audra, and C. Brereton. Notre Dame: University of Notre Dame Press.

Bergström, Lars. 1996. Reflections on Consequentialism. *Theoria* 62: 74–94.

Blackburn, Simon. 1996 (1994). Foundationalism. In *The Oxford Dictionary of Philosophy*. Oxford: Oxford University Press.

Brandt, Richard. 1958. Blameworthiness and Obligation. In *Essays in Moral Philosophy*, ed. A.I. Melden. Seattle: University of Washington Press.

———. 1959. *Ethical Theory: The Problems of Normative and Critical Ethics*. Englewood Cliffs: Prentice-Hall.

Bricke, John, ed. 1976. *Freedom & Morality*. Lawrence: University of Kansas.

J. Österberg, *Towards Reunion in Ethics*, Philosophical Studies Series 138,
https://doi.org/10.1007/978-3-030-12410-6

Brink, David. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
Brinton, Crane. 1959. *A History of Western Morals*. New York: Harcourt/Brace and Company.
Broad, C.D. 1930. *Five Types of Ethical Theory*. London: Routledge & Kegan Paul.
———. 1942. Certain Features in G. E. Moore's Ethical Doctrines. In *The Philosophy of G. E. Moore*, ed. Paul Schilpp, 43–67. Chicago/Evanston: Northwestern University Press.
Brock, Dan. 1973. Recent Work in Utilitarianism. *American Philosophical Quarterly* 10: 241–276.
Brown, Thomas. 1820. *Lectures on the Philosophy of the Human Mind*, 4 vols. Edinburgh: Tait.
Brown, Stuart M., Jr., ed. 1950. *Five Sermons Preached at the Rolls Chapel and a Dissertation Upon the Nature of Virtue*. Indianapolis: Bobbs-Merrill.
Buchanan, James. 1975. *The Limits of Liberty: Between Anarchy and Leviathan*. Chicago: Chicago University Press.
Butler, Joseph. 1726. *Fifteen Sermons Preached at the Rolls Chapel*. London: James and John Knapton.
Bykvist, Krister. 2003. Normative Supervenience and Consequentialism. *Utilitas* 15: 27–49.
Carlson, Erik. 1995. *Consequentialism Reconsidered*. Dordrecht: Kluwer.
Chan, Wing-tsit, ed. 1963. *A Source Book in Chinese Philosophy*. Princeton: Princeton University Press.
Clarke, Samuel. 1987. A Discourse Concerning the Unchangeable Obligations of Natural Religion, and the Truth and Certainty of the Christian Revelation. 2nd series of the Boyle lectures, delivered at St. Paul's in 1705; quoted from Mackie, J.L. 1980. *Hume's Moral Theory*. London: Routledge and Kegan Paul.
Collins Dictionary of the English Language. Glasgow: Collins, 1985.
Commins, Saxe, and Robert Linscott, (eds.). 1947. *The World's Great Thinkers*, Vol. 3: *Man and the State.* Trans. Samuel Moore. New York: Random House.
Cooper, Wesley E. et al. (eds.). 1979. New Essays on John Stuart Mill and Utilitarianism. *Canadian Journal of Philosophy*, Suppl. Vol. V.
Copp, David, ed. 2006. *The Oxford Handbook of Ethical Theory*. New York: Oxford University Press.
Crisp, Roger. 1997. *Mill on Utilitarianism*. London: Routledge.
Crisp, Roger, and Michael Slote, eds. 1997. *Virtue Ethics*. Oxford: Oxford University Press.
Cummiskey, David. 1990. Kantian Consequentialism. *Ethics* 100: 586–615.
———. 1996. *Kantian Consequentialism*. New York/Oxford: Oxford University Press.
Dancy, Jonathan. 1993. *Moral Reasons*. Oxford: Blackwell.
———. 1998. Moral realism. In *Routledge Encyclopedia of Philosophy*, vol. 6, 534–539. London/New York: Routledge.
Daniels, Norman. 1979. Wide Reflective Equilibrium and Theory Acceptance in Ethics. *The Journal of Philosophy* 76: 256–282.
———. 1980. Reflective Equilibrium and Archimedean Points. *Canadian Journal of Philosophy* 10: 83–103.
Danielsson, Sven. 1988. Konsekvensetikens gränser (The Limits of Consequentialism). In *Filosofiska utredninga*. N.p.: Thales.
———. 1998. *Filosofiska utredningar*. N.p.: Thales.
Danto, Arthur. 1965. Basic Actions. *American Philosophical Quarterly* 2: 141–148.
———. 1976. *Mysticism and Morality: Oriental Thought and Moral Philosophy*. Harmondsworth: Penguin.
de Waal, Fran. 2006. *Primates and Philosophers: How Morality Evolved*. Eds. and Intr. Stephen Macedo and Josiah Ober. Princeton/Oxford: Oxford University Press/Princeton University Press.
DePaul, Michael. 1993. *Balance and refinement: Beyond coherence methods of moral inquiry*. London/New York: Routledge.
Dickie, George. 1984. The New Institutional Theory of Art. In *Proceedings of the Eighth International Wittgenstein Symposium, Part I*, ed. Rudolf Haller, 57–64. Vienna: Holder-Pichler-Temsky.

Dickie, George, et al., eds. 1977. *Aesthetics: A Critical Anthology*. 2nd ed. New York: St. Martin's Press.
Doberstein, John. 1959. *Luther's Works*. Trans. and ed. Philadelphia: Fortress.
Donagan, Alan. 1977. *The Theory of Morality*. Chicago: University of Chicago Press.
Edwards, Paul, ed. 1967. *The Encyclopedia of Philosophy*. New York: Macmillan.
Eriksson, Björn. 1994. *Heavy Duty: On the Demands of Consequentialism*. Stockholm: Almqvist & Wiksell International.
Ezorsky, Gertrude. 1974. Unconscious Utilitarianism. *The Monist* 58: 468–474.
Falk, W.D. 1986. Morality, Self, and Others. In *Ought, Reasons, and Morality: The Collected Papers of W. D. Falk*, 198–231. Ithaca: Cornell University Press.
Firth, Roderick. 1952. Ethical Absolutism and the Ideal Observer. *Philosophy and Phenomenological Research* 12: 317–345.
Fischer, John Martin, and Mark Ravizza, eds. 1992. *Ethics: Problems and Principles*. Fort Worth: Harcourt Brace Jovanovich.
Fletcher, Joseph. 1966. *Situation Ethics*. Philadelphia: The Westminster Press.
Flusser, David. 1990. The Ten Commandments and the New Testament. In *The Ten Commandments in History and Tradition*, ed. Gershon Levi. Jerusalem: The Magnes Press/The Hebrew University. (1985).
Foot, Philippa. 1988. Utilitarianism and the Virtues. In *Consequentialism and Its Critics*, ed. Samuel Scheffler, 224–242. Oxford: Oxford University Press. Originally published in *Mind* 94 (1985): 196–209.
Frankena, William. 1973. *Ethics*. 2nd ed. Englewwod Cliffs: Prentice-Hall.
———. 1990. Hare on Levels of Moral Thinking. In *Hare and Critics: Essays on Moral Thinking*, ed. Douglas Seanor and N. Fotion. Oxford: Clarendon Press.
Freud, Sigmund. 1953–1964. *Civilization and Its Discontents*. First published in 1930, it is included in Vol. XXI of *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: The Hogarth Press.
Fried, Charles. 1978. *Right and Wrong*. Cambridge: Harvard University Press.
Fromm, Erich. 1947. *Man for Himself: An Inquiry into the Psychology of Ethics*. New York: Rinehart.
Galling, Kurt, (ed.). 1958. Goldene Regel. In *Die Religion in Geschichte und Gegenwart: Handwörterbuch für Theologie und Religionswissenschaft*. Tübingen: J. C. B. Mohr (Paul Siebeck).
Garnett, A.C. 1960. *Ethics: A Critical Introduction*. New York: Ronald Press.
Gauthier, David, ed. 1970. *Morality and Rational Self-Interest*. Englewood Cliffs: Prentice-Hall.
———. 1975. Coordination. *Dialogue: Canadian Philosophical Review* 14: 195–221.
———. 1979. David Hume: Contractarian. *Philosophical Review* 88: 3–38.
———. 1986. *Morals by Agreement*. Oxford: Oxford University Press.
———. 1990. *Moral Dealing: Contract, Ethics, and Reason*. Ithaca/London: Cornell University Press.
Geach, Peter. 1956. Good and Evil. *Analysis* 17: 33–42.
Gensler, Harry. 1996. *Formal Ethics*. London/New York: Routledge.
Gert, Bernard. 2005 (1988). *Morality: Its Nature and Justification*. New York: Oxford University Press.
Gewirth, Alan. 1976. Moral Rationality. In *Freedom & Morality*, ed. John Bricke, 113–150. Lawrence: University of Kansas.
———. 1978. *Reason and morality*. Chicago: Chicago University Press.
Gibbard, Allan. 1978. Act-Utilitarian Agreements. In *Values and Morals*, ed. A.I. Goldman and J. Kim, 91–119. Dordrecht: Reidel.
———. 1982. Inchoately Utilitarian Common Sense: The Bearing of a Thesis of Sidgwick's on Moral Theory. In *The Limits of Utilitarianis*, ed. B.Miller Harlan and William H. Williams. Minneapolis: University of Minnesota Press.

———. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Oxford: Oxford University Press.
Glossop, Ronald. 1976. Is Hume a 'Classical Utilitarian'. In *Hume Studies*, vol. 2, 1–16.
Glover, Jonathan. 1975. It Makes No Difference Whether or Not I do It. *Proceedings of the Aristotelian Society* 49 (Supp): 171–190.
Godwin, William. 1793. *Enquiry Concerning Political Justice*. Dublin: Luke White.
Goldman, A.I., and J. Kim, eds. 1978. *Values and Morals*. Dordrecht: Reidel.
Green, Thomas. 1798. *An Examination of the Leading Principle in the New System of Morals*
Greene, Joshua D., et al. 2001. An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science* 293: 2105–2108.
Griffin, James. 1986. *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Oxford University Press.
Griffiths, Phillips. 1957/58. Justifying Moral Principles. *Proceedings of the Aristotelian Society, N. S.* 58: 103–124.
———. 1967. Ultimate Moral Principles: Their Justification. In *The Encyclopedia of Philosophy*, ed. Paul Edwards, vol. 8, 177–182. New York: Macmillan.
Grote, John. 1870. *An Examination of the Utilitarian Philosophy*. Cambridge: Deighton, Bell & Co.
Halévy, Elie. 1901. *La formation du radicalisme philosophique*. Paris: Alcan. Trans. Mary Morris as *The Growth of Philosophical Radicalism*. London: Faber and Gwyer, 1928.
Haller, Rudolf, ed. 1984. *Proceedings of the Eighth International Wittgenstein Symposium, Part I*. Vienna: Holder-Pichler-Temsky.
Hardin, Russell. 1988. *Morality Within the Limits of Reason*. Chicago/London: The University of Chicago Press.
Hare, R.M. 1963. *Freedom and Reason*. Oxford: Oxford University Press.
———. 1975a. Abortion and the Golden Rule. *Philosophy and Public Affairs* 3: 201–222.
———. 1975b. Euthanasia: A Christian View. *Proceedings of the Center for Philosophic Exchange* 6: 43–52.
———. 1976. Ethical Theory and Utilitarianism. In *Contemporary British Philosophy*, ed. H.D. Lewis, 113–131. London: Allen and Unwin.
———. 1981. *Moral Thinking: Its Levels, Methods, and Point*. Oxford: Oxford University Press.
———. 1990. Comments. In *Hare and Critics: Essays on Moral Thinking*, ed. Douglas Seanor and N. Fotion. Oxford: Clarendon Press.
———. 1993. Could Kant Have Been a Utilitarian. *Utilitas* 5: 1–16.
Harman, Gilbert. 1977. *The Nature of Morality: An Introduction to Ethics*. New York: Oxford University Press.
Harman, Gilbert, and Judith Jarvis Thomson. 1996. *Moral Relativism and Moral Objectivity*. Oxford: Blackwell.
Harris, N.G.E. 1972. Nondeliberative Utilitarianism. *Ethics* 82: 344–348.
Harris, John. 1980. *Violence and Responsibility*. London: Routledge and Kegan Paul.
Harsanyi, John. 1977a. Morality and the Theory of Rational Behavior. *Social Research* 44: 623–656.
———. 1977b. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.
Hart, H.L.A. 1961. *The Concept of Law*. Oxford: Oxford University Press.
Hartnack, Justus. Performative Utterances. In *The Encyclopedia of Philosophy*, ed Paul Edwards.
Hastings, James, ed. 1908–1926. *Encyclopedia of Religion and Ethics*, 13 vols. Edinburgh/New York: T. and T. Clark/Charles Scribner's sons.
Epstein, I., ed. 1987. *Hebrew-English Edition of the Babylonian Talmud*. London: Soncino, Cop.
Hegel G.W.F. 1991. *Elements of the Philosophy of Right*. Trans. H.B. Nisbet, ed. Allen W. Wood. Cambridge: Cambridge University Press.
Henry, Sidgwick. 1907. *The Methods Of Ethics*. 7th ed. London: Macmillan.

Herodotus. 1963. *Histories*. 3:38; The Loeb classical library, ed. and tr. A.D. Godley, Vol. 2. Cambridge, MA: Heineman (1921).
Hodgson, D.H. 1967. *Consequences of Utilitarianism: A Study in Normative Ethics and Legal Theory*. Oxford: Oxford University Press.
Hoerster, Norbert. 1973. Is Act-Utilitarian Truth-Telling Self-Defeating? *Mind* 82: 413–416.
Holmes, Jr, and Wendell Oliver. 1896. The Path of the Law. *Harvard Law Review* 10: 457–478.
Hooker, Brad. 2002 (2000). *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Oxford: Oxford University Press, .
Hooker, Brad. Rule-Consequentialism. In *The Stanford Encyclopedia of Philosophy (Spring 2004 Edition)*, ed. Edward N. Zachta. http://plato.stanford.edu/archives/spr2004/entries/consequentialism/rule, p. 12.
Hume, David. 1998. *An Enquiry concerning the Principles of Morals*. Oxford: Oxford University Press.
———. 2004 (2000). *A Treatise of Human Nature*, ed. David Fate Norton and Mary Fate Norton. Oxford: Oxford University Press.
Hurka, Thomas. 1990. Two Kinds of Satisficing. *Philosophical Studies* 59: 107–111.
Hursthouse, Rosalind. 1999. *On Virtue Ethics*. Oxford: Oxford University Press.
Hutcheson, Francis. 1742. *An Essay on the Nature and Conduct of the Passions and Affections* and *Illustrations upon the Moral Sense.* 3rd ed. London: printed for A. Ward et al. (1728).
Isocrates. 1928. Nicocles. In *Loeb Classical Library*. Cambridge: Harvard University Press.
Jackson, Samuel M., ed. 1949. *The New Schaff-Herzog Encyclopedia of Religius Knowledge*. Grand Rapids: Baher Book House.
Jackson, Frank. 1991. Decision-Theoretic Consequentialism and the Nearest and Dearest Objection. *Ethics* 101: 461–482.
Jarvis, Thomson Judith, and Gerald Dworkin, eds. 1968. *Ethics*. New York: Harper & Row.
Jeffrey, Wattles. 1996. *The Golden Rule*. Oxford: Oxford University Press.
Joel, Feinberg, ed. 1969. *Moral Concepts*. London: Oxford University Press.
Kagan, Shelly. 1989. *The Limits of Morality*. Oxford: Oxford University Press.
———. 1991. Précis of The Limits of Morality. *Philosophy and Phenomenological Research* 51: 897–901.
———. 1998. *Normative Ethics*. Boulder: Westview Press.
Kalin, Jesse. 1970. In Defense of Egoism. In *Morality and Rational Self-Interest*, ed. David Gauthier, 64–87. Englewood Cliffs: Prentice-Hall.
Kant, Immanuel. 1873. *Kritik der praktischen Vernunft*; quoted from *Kant's Critique of Practical Reason and Other Works on the Theory of Ethics*. Trans. T.K. Abbott. London: Longmans, Green.
———. 1948. *Grundlegung zur Metaphysik der Sitten*, 404; the quotation is from H. J. Paton's translation, *The Moral Law.* London: Hutchinson.
———. 1976. On a Supposed Right to Lie from Altruistic Motives. In *Immanuel Kant: Critique of Practical Reason and Other Writings in Moral Philosophy*, edited and translated by Lewis White Beck, 346–350. Reprint. New York: Garland.
Kapur, Neera. 1991. Why Is It Wrong to Be Always Guided by the Best: Consequentialism and Friendship. *Ethics* 101: 483–504.
Keynes, John Maynard. 1949. My Early Beliefs. In *Two Memoirs*. New York: Augustus M. Kelley.
Knud, Haakonssen. 2002. Article on Mackintosh. In *Dictionary of Nineteenth-Century British Philosophers*, ed. W.J. Mander and Alan P.F. Sell, vol. 2, 715–719. Bristol: Thoemmes Press.
Korsgaard, Christine. 1998. Teleological Ethics. In *Routledge Encyclopedia of Philosophy*. London/New York: Routledge.
Kuflik, Arthur. 1986. A Defense of Common-Sense Morality. *Ethics* 96: 784–803.
Kupperman, Joel. 1983. *The Foundations of Morality*. London: George Allen & Unwin.
Kymlicka, Will. 1988. Rawls on Teleology and Deontology. *Philosophy & Public Affairs* 17: 173–190.

———. 1993. The Social Contract Tradition. In *A Companion to Ethics*, ed. Peter Singer, 186–196. Oxford: Blackwell.
Leibniz, G.W. 1981. *New Essays on Human Understanding*. Trans. Peter Remnant and Jonathan Bennett. Cambridge: Cambridge University Press (1765).
Levi, Gershon (ed.). 1990 (1985). The Ten Commandments in History and Tradition. Jerusalem: The Magnes Press, The Hebrew University.
Lewis, David. 1969. *Convention: A Philosophical Study*. Cambridge: Harvard University Press.
———. 1972. Utilitarianism and Truthfulness. *Australasian Journal of Philosophy* 50: 17–19.
Luce, Duncan, and Howard Raiffa. 1957. *Games and Decisions*. New York: Wiley.
Lukes, Steven. 1973. *Individualism*. Oxford: Blackwell.
Lyons, David. 1965. *Forms and Limits of Utilitarianism*, 1970. Oxford: Oxford University Press.
Mackie, J.L. 1973. The Disutility of Act-Utilitarianism. *The Philosophical Quarterly* 23: 289–300.
———. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin.
———. 1987. *Hume's Moral Theory*. London: Routledge and Kegan Paul. (1980).
Mackintosh, James. 1836. *A Dissertation on the Progress of Ethical Philosophy, Chiefly During the Seventeenth and Eighteenth Centuries*. Edinburgh: Adam and Charles Black.
Mander, W.J. and Alan P.F. Sell (eds.). 2002. Dictionary of Nineteenth-Century British Philosophers, 2 vols. Bristol: Thoemmes Press.
Marx, Karl. 1947. The Communist Manifesto. In *The World's Great Thinkers*, ed Saxe Commins and Robert Linscott, Vol. 3: *Man and the State;* translated by Samuel Moore. New York: Random House.
———. 1977. The German Ideology. In *Karl Marx: Selected writings*, ed. David McLellan. Oxford: Oxford University Press.
Mautner, Thomas, ed. 1997. *Dictionary of Philosophy*. 2nd ed. Harmondsworth: Penguin.
McCloskey, H.J. 1965. A Non-Utilitarian Approach to Punishment. *Inquiry* 8: 249–263.
McLellan, David, ed. 1977. *Karl Marx: Selected Writings*. Oxford: Oxford University Press.
McNaughton, David, and Piers Rawling. 2006. Deontology. In *The Oxford Handbook of Ethical Theory*, ed. David Copp, 424–458. New York: Oxford University Press.
Melden, A.I., ed. 1958. *Essays in Moral Philosophy*. Seattle: University of Washington Press.
Midgley, Mary. 1993. The Origin of Ethics. In *A Companion to Ethics*, ed. Peter Singer, 3–13. Oxford: Blackwell.
Mill, J.S. 1998. In *Utilitarianism*, ed. Roger Crisp. Oxford: Oxford University Press.
Miller, Harlan B., and William H. Williams, eds. 1982. *The Limits of Utilitarianism*. Minneapolis: University of Minnesota Press.
Monro, D.H., ed. 1972. *A Guide to the British Moralists*. London: Collins.
Moore, G.E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.
———. 1912. *Ethics*. London: Oxford University Press.
Moore, G. E. 1942. Autobiography. In *The Philosophy of G. E. Moore*, The Library of Living Philosophers, Vol. IV, ed. Paul A. Schilpp. Evanston: Northwestern University Press.
Muirhead, J.H. 1932. *Rule and End in Morals*. London: Oxford University Press.
Myers, R.H. 1994. Prerogatives and Restrictions from the Cooperative Point of View. *Ethics* 105: 128–152.
Nagel, Thomas. 1970. *The Possibility of Altruism*. Oxford: Oxford University Press.
———. 1979a. Moral Luck. *P. A. S.*, Supp. Vol. L (1976): 137–155; Reprinted in *Mortal Questions*. Cambridge: Cambridge University Press.
———. 1979b. *Mortal Questions*. Cambridge: Cambridge University Press.
———. 1986. *The View from Nowhere*. Oxford: Oxford University Press.
Nansen, Fridtjof. 1890. *The First Crossing of Greenland*, 2 vols. London: Longman/Green & Co.
Narveson, Jan. 1971. Promising, Expecting, and Utility. *Canadian Journal of Philosophy* 1: 207–223.
Nell, Onora. 1975. *Acting on Principle: An Essay on Kantian Ethics*. New York/London: Columbia University Press.
Niebuhr, Reinhold. 1949. *The Nature and Destiny of Man*. New York: Charles Scribner's Sons.

Norman, Richard. 1983. *The Moral Philosophers: An Introduction to Ethics*. Oxford: Oxford University Press.
Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge: Harvard University Press.
Olson, Robert. 1967a. Deontological Ethics. In *The Encyclopedia of Philosophy*, ed. Paul Edwards. New York: Macmillan.
———. 1967b. Teleological Ethics. In *The Encyclopedia of Philosophy*, ed. Paul Edwards. New York: Macmillan.
Olson, Jonas, and Frans Svenson. 2003. A Particular Consequentialism. *Utilitas* 15: 194–205.
Österberg, Jan. 1988. *Self and Others: A Study of Ethical Egoism*. Dordrecht: Kluwer.
———. 1999. The Virtues of Virtue Ethics. In *Philosophical crumbs*, ed. Rysiek Sliwinski. Uppsala: Uppsala Philosophical Studies.
Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
———. 1988. Is Common-Sense Morality Self-Defeating? In *Consequentialism and Its Critics*, ed. Samuel Scheffler, 173–186. Oxford: Oxford University Press. Originally published in *Journal of Philosophy*, 76 (1979): 533–45.
———. Draft of 28 April 2008. *On what matters*. Derek Parfit's Homepage.
Philippidis, L.J. 1929. *Die "goldene Regel" religionsgeschichtlich untersucht*. Leipzig: Adolf Klein Verlag.
Piper, Adrian. 1978. Utility, Publicity, and Manipulation. *Ethics* 88: 189–206.
Plato. 1987. *The Republic*. Trans Desmond Lee. Harmondsworth: Penguin.
Postow, B.C. 1977. Generalized Act Utilitarianism. *Analysis* 37: 49–52.
Provis, C. 1977. Gauthier on Coordination. *Dialogue: Canadian Philosophical Review* 16: 507–509.
Rabinowicz, Wlodek, and Jan Österberg. 1996. Value Based on Preferences: On Two Interpretations of Preference Utilitarianism. *Economics and Philosophy* 12: 1–27.
Rachels, James. 1974. Two Arguments against Ethical Egoism. *Philosophia* 4: 297–314.
———. 1999. *The Elements of Moral Philosophy*. 3rd ed. Boston: McGraw-Hill.
Railton, Peter. 1988. Alienation, Consequentialism, and the Demands of Morality. In *Consequentialism and Its Critics*, ed. Samuel Scheffler, 93–133. Oxford: Oxford University Press. Originally published in *Philosophy and Public Affairs* 13 (1984): 134–71.
Raphael, D.D. 1969. *British Moralists 1650–1800*, 2 vols. Oxford: Oxford University Press.
———. 1974. Sidgwick on Intuitionism. *The Monist* 58: 405–419.
Rawls, John. 1971. *A Theory of Justice*. Cambridge: Harvard University Press.
Regan, Donald. 1984 (1980). Utilitarianism and Cooperation. Oxford: Oxford University Press.
Regan, Tom. 1986. *Bloomsbury's Prophet: G. E. Moore and the Development of His Moral Philosophy*. Philadelphia: Temple University Press.
Reich, Wilhelm. 1969. *The Sexual Revolution: Toward a Self-Governing Character Structure*. New York: Farrar, Straus and Giroux.
Reiner, Hans. 1983. *Duty and Inclination: The Fundamentals of Morality Discussed and Redefined with Special Regard to Kant and Schiller*. The Hague: Martinus Nijhoff.
Rescher, Nicholas. 1966. *Distributive Justice: A Constructive Critique of the Utilitarian Theory of Distribution*. Indianapolis: Bobbs-Merrill.
Richards, David. 1970. *A Theory of Reasons for Action*. Oxford: Oxford University Press.
Ross, W.D. 1930. *The Right and the Good*. London: Oxford University Press.
———. 1939. *Foundations of Ethics*. London: Oxford University Press.
Rousseau, Jean-Jacques. 1973. *The Social Contract and Discourses*. Trans G.D.H. Cole. London: Everyman.
Routledge Encyclopedia of Philosophy. London/New York: Routledge, 1998.
Runciman, W.G., and Amartya Sen. 1965. Games, Justice and the General Will. *Mind* 74: 554–562.
Rydén, Lars (ed.). 1990. *Etik för forskare: en antologi med utgångspunkt i arbetet med Uppsalakoden* ("Ethics for Scientists: An Anthology with a Point of Departure in the Working with the Uppsala Code"). Stockholm: UHÄ/foU skriftserie, 1990: 1.

Scanlon, T.M. 1988. Levels of Moral Thinking. In *Hare and Critics: Essays on Moral Thinkin*, ed. Douglas Seanor and N. Fotion. Oxford: Oxford University Press. (Oxford: Clarendon Press, 1990).
———. 1998. *What we owe to each other*. Cambridge/London: Harvard University Press.
Scarre, Geoffrey. 1996. *Utilitarianism*. London/New York: Routledge.
Scheffler, Samuel, ed. 1988a. *Consequentialism and Its Critics*. Oxford: Oxford University Press.
———. 1988b. Introduction. In *Consequentialism and Its Critics*, ed. Samuel Scheffler. Oxford: Oxford University Press.
———. 2000 (1982). *The Rejection of Consequentialism*. Oxford: Oxford University Press.
Schilpp, Paul, ed. 1942. *The Philosophy of G. E. Moore*. Chicago/Evanston: Northwestern University Press.
Schneewind, J.B. 1963. First Principles and Common Sense Morality in Sidgwick's Ethics. *Archiv für Geschichte der Philosophie, Bd.* 45: 137–156.
———. 1977. *Sidgwick's Ethics and Victorian Moral Philosophy*. Oxford: Oxford University Press.
———. 1990. The Misfortunes of Virtue Ethics. *Ethics* 101: 42–63.
Seanor, Douglas, and N. Fotion, eds. 1988. *Hare and Critics: Essays on Moral Thinking*. Oxford: Oxford University Press.
Sedgwick, Adam. 1832. *Discourse on the Studies of the University of Cambridge*. Cambridge: Cambridge University Press.
Selby-Bigge, L.A., ed. 1897. *British Moralists*. Oxford: Oxford University Press.
Sen, Amartya. 1970. *Collective Choice and Social Welfare*. San Francisco/Edinburgh: Holden-Day/Oliver & Boyd.
Sen, Amartya, and Bernard Williams, eds. 1982. *Utilitarianism and Beyond*. Cambridge: Cambridge University Press.
Shakespeare *Hamlet*. Act 5, Scene 2, lines 129–132.
Shaw, George Bernard. 1919 (1903). Maxims for a Revolutionist. In *Man and Superman*. London: Constable.
Simmons, John. 1982. Utilitarianism and Unconscious Utilitarianism. In *The Limits of Utilitarianism*, ed. Harlan B. Miller and William H. Williams, 86–92. Minneapolis: University of Minnesota Press.
Simon, Herbert. 1959. Theories of Decision Making in Economics and Behavioral Science. *American Economic Review* 49: 253–283.
Singer, Marcus G. 1967. Golden Rule. In *The Encyclopedia of Philosophy*, ed. Paul Edwards, vol. 3, 365–367. New York: Macmillan.
Singer, Peter. 1972. Is Act-Utilitarianism Self-Defeating? *Philosophical Review* 81: 94–104.
———. 1974. Sidgwick and Reflective Equilibrium. *The Monist* 58: 490–516.
———. 1981. *The Expanding Circle: Ethics and Sociobiology*. Oxford: Oxford University Press.
———, ed. 1993. *A Companion to Ethics*. Oxford: Blackwell.
———. 2005. Ethics and Intuitions. *The Journal of Ethics* 9: 331–352.
Slote, Michael. 1985. *Common-sense Morality and Consequentialism*. London: Routledge & Kegan Paul.
———. 1997. Virtue Ethics. In *Three Methods of Ethics: A Debate*, ed. Marcia Baron, Philip Pettit, and Michael Slote. Oxford: Blackwell.
———. 2001. *Moral from Motives*. Oxford: Oxford University Press.
Smart, J.J.C. 1956. Extreme and Restricted Utilitarianism. *The Philosophical Quarterly* 6: 344–354.
———. 1973. An Outline of a System of Utilitarian Ethics. In *Utilitarianism: For and Against* by J.J. C. Smart and Bernard Williams, Ch. 1. London: Cambridge University Press.
Smart, J.J.C., and Bernard Williams. 1973. *Utilitarianism: For and Against*. London: Cambridge University Press.
Sobel, Jordan Howard. 1989. Kent Bach on Good Arguments. *Canadian Journal of Philosophy* 19: 447–454.

———. 2008. *Walls and Vaults*. New York: Wiley.
Spooner, W.A. 1913. The Golden Rule. In *Encyclopedia of Religion and Ethics*, ed. James Hastings. Edinburgh: T. & T. Clark.
Stephen, Leslie. 1900. *The English Utilitarians*. London: Duckworth.
Strang, Colin. 1960. What If Everyone Did That? *Durham University Journal* 53: 5–10.
Strawson, P.F. 1961. Social Morality and Individual Ideal. *Philosophy* 36: 1–17.
———. 1970. Social Morality and Individual Ideal. In *The Definition of Morality*, ed. G. Wallace and A.D.M. Walker, 98–118. London: Methuen.
Sumner, L.W. 1969. Consequences of Utilitarianism. *Dialogue: Canadian Philosophical Review* 7: 639–642.
Svensson, Frans. 2006. *Some Basic Issues in Neo-Aristotelian Virtue Ethics*. Uppsala: n. p.
Thieme, Karl. 1949. Consilia Evangelica. In *The New Schaff-Herzog Encyclopedia of Religius Knowledge*, ed. Samuel M. Jackson. Grand Rapids: Baher Book House.
Hobbes Thomas. 1968. Leviathan, ed. C.B. MacPherson. Harmondsworth: Penguin.
Thomson, Judith Jarvis. 1985. The Trolley Problem. *The Yale Law Journal* 94: 1395–1415. Reprinted in *Ethics: Problems and Principles* edited by Fischer, John Martin and Mark Ravizza, 279–292. Fort Worth: Harcourt Brace Jovanovich, 1992.
———. 1997. The Right and the Good. *The Journal of Philosophy* 94: 273–298.
Tiberius, Valerie. 2006. How to Think About Virtue and Right. *Philosophical Papers* 35: 247–265.
Turnbull, Colin. 1972. *The Mountain People*. New York: Simon and Schuster.
Unger, Peter. 1996. *Living High and Letting Die: Our Illusion of Innocence*. Oxford: Oxford University Press.
Urmson, J.O. 1953. The Interpretation of the Moral Philosophy of J. S. Mill. *The Philosophical Quarterly* 3: 33–40.
———. 1969. Saints and Heroes. In *Moral Concepts*, ed. Joel Feinberg, 60–73. London: Oxford University Press. Originally published in *Essays in Moral Philosophy* edited byA. I. Melden, 198–216. Washington: Washington University Press, 1958.
von Wright, G.H. 1963. *The Varieties of Goodness*. London: Routledge & Kegan Paul.
Waldron, Jeremy. 1994. Kagan on Requirements: Mill on Sanctions. *Ethics* 104: 310–324.
Wallace, G., and A.D.M. Walker, eds. 1970. *The Definition of Morality*. London: Methuen.
Warnock, G.J. 1976 (1971). *The Object of Morality*. London: Methuen.
West, Henry R. 2004. *An Introduction to Mill's Utilitarian Ethics*. Cambridge: Cambridge University Press.
Westermarck, Edward. 1908. *The Origin and Development of the Moral Ideas*, 2 vols. London: Macmillan.
Whewell, William. 1836. Preface to *A Dissertation on the Progress of Ethical Philosophy, Chiefly During the Seventeenth and Eighteenth Centuries* by James Mackintosh. Edinburgh: Adam and Charles Black.
William, Bishop of St. Davids. 1679. *The Comprehensive Rule of Righteousness: Do as You Would Be Done By*. Cornhill: William Leach.
Williams, Bernard. "A Critique of Utilitarianism". In Utilitarianism: For and Against, by J. J. C. Smart and Bernard Williams. London: Cambridge University Press, 1973.
———. 1981a. Moral Luck. *P. A. S.* Supp. Vol. L (1976): 115–135. Reprinted in *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press.
———. 1981b. Utilitarianism and Moral Self-Indulgence. In *Moral Luck*. Cambridge: Cambridge University Press.
———. 1981c. *Moral Luck*. Cambridge: Cambridge University Press.
Wood, Allen W. 1990. *Hegel's Ethical Thought*. Cambridge: Cambridge University Press.
Zuboff, Arnold. 1977–1978. Moment Universals and Personal Identity. *Proceedings of the Aristotelian Society N. S* 77: 141–155.
———. 1990. One Self: The Logic of Experience. *Inquiry* 33: 39–68.