

## Alternatives\*

Gunnar Björnsson  
University of Gothenburg

### Introduction

- (1) What are *the* alternatives in a situation of deliberation and planning?
- (2) What should a coherent formulation of act-utilitarianism look like?

These two questions have been at the focus of interest in a debate initiated by Lars Bergström's thesis *The Alternatives and Consequences of Action*. In this paper I wish to examine some of the proposed answers to (1) and (2), and some of the arguments given in favour of these answers. I will concentrate on the part of the debate represented in the literature listed at the end of the essay.

It is hard to present the alleged problems and the proposed solutions in a way that does justice to the participants of the debate without actually letting them speak for themselves, through their own papers. I will nevertheless try to give the reader some feeling of what is at stake without indulging in *all* the technical subtleties of the discussion. More than simply expose the ideas of others, I intend to discuss what's sound and unsound in the ideas put forward, and I am also presenting my own position at the end of the paper.

In part one I present the theories of the debaters and apply them to two imaginary cases in order to highlight their differences. In part two I propose and discuss two criteria of adequacy for utilitarian principles. In part three, the criteria are applied to one of the imaginary cases from part one, and used in a critical discussion of the theories presented there. In part four finally, I outline my own theory.

Even though the debate displays some fundamental disagreements, it is best understood against a background of agreement. Thus every participant accepts – at least for the sake of argument – what could be called ‘the spirit of act-utilitarianism’. The classic formulation of one of the most fundamental principles of utilitarianism is the following:

- (U) An action, *a*, ought to be performed if and only if its consequences are better than those of every alternative to *a*.

---

\* This text was written almost twenty years ago and discussed in Erik Carlson's *Consequentialism Reconsidered* (Dordrecht: Kluwer 1995). Since some people have asked for copies, I decided to transfer it to electronic form. I've tried to purge the text of numerous inevitable scanning errors, but readers beware!

According to most utilitarians the concept of ‘ought’ is relative to a scale of value on which the (total) consequences of actions are compared. It is also taken to be relative to a set of alternatives, including  $a$  and every alternative to  $a$ . Such a set is for obvious reasons called an *alternative-set*.

The following restrictions on alternative-sets to which (U) applies were proposed and argued for by Bergström in [1] (p 29 ff):

- (AS) The set  $A$  is an alternative-set if, and only if, (i) every member of  $A$  is a particular action<sup>1</sup>, (ii)  $A$  has at least two different members, the members of  $A$  are (iii) agent-identical<sup>2</sup>, (iv) time-identical<sup>3</sup>, (v) performable, (vi) incompatible in pairs<sup>4</sup>, and (vii) jointly exhaustive<sup>5</sup>. If  $a$  and  $b$  are alternatives, they are members of the same alternative-set.

‘Every alternative’ in (U) should then be understood as ‘all members of the same alternative-set as the action in question’. The restrictions seem to be reasonable. But this specification of ‘every alternative’ seems to give the utilitarian severe problems in most or all cases. The underlying reason is that it seems that *for a given person and a given situation, there are several alternative-sets*, that is: several sets of alternatives fulfilling the conditions in (AS). I will now look at such a situation.

### Case One

A person,  $P$ , at a time,  $t$ , can perform each of and only the following actions:  $a$  (have a cigarette),  $b$  (not have a cigarette),  $a'$  (have a cigarette and read a paper on smoking and lung-cancer) and  $a''$  (have a cigarette, but read no paper on smoking and lung-cancer). The consequences of these actions are valued as follows:  $a'$  has better consequences than  $b$  while  $a$  and  $a''$  has worse consequences than both  $a'$  and  $b$ .

Assuming that these actions are all time-identical,  $\{a, b\}$  as well as  $\{a', a'', b\}$  seem to be alternative-sets according to (AS). If (U) is applied to  $\{a, b\}$ , the result seems to be that  $b$  ought to be done, while  $a$  ought not to be done. If, on the other hand, (U) is applied to  $\{a', a'', b\}$  the result is that  $a'$  ought to be done, while  $a''$  and  $b$  ought not to be done. We are now facing two apparent problems:

- (P1) In Case One,  $a'$  ought to be done, and if  $a'$  is done this implies that  $a$  is done, but  $a$  ought not to be done in Case One.

---

<sup>1</sup> A particular action is relative both to persons and times. My typing now is for example a particular action. Particular actions could be contrasted with generic actions, which are not relative to both persons and times. “My typing”, “typing now” and “typing” are all generic actions.

<sup>2</sup> Two actions are agent-identical if they have the same agent.

<sup>3</sup> Two actions are time-identical if they have the same time-interval.

<sup>4</sup> This restriction is invoked to make sure that if one of the alternatives is performed, none of the other alternatives is performed.

<sup>5</sup> If the members of  $A$  is jointly exhaustive, at least one of the members of  $A$  must be performed. Together with (vi), (vii) ensures that one and only one of the members of an alternative-set will ever be performed.

- (P2) In Case One, it is true that  $b$  ought to be done, but it is also true that  $b$  ought not to be done.

(P2) is an outright paradox. (P1) would be paradoxical if the following principle is true:

- (c) If an action  $a'$  ought to be done, and if that  $a'$  is done logically implies that  $a$  is done, then  $a$  ought to be done.

Maybe (c) is intuitively plausible. With (c), (P1) implies that  $a$  ought to be done and that  $a$  ought not to be done. This is surely a paradox as it stands. The situation can also be seen as a moral dilemma – if P does  $a'$ , which ought to be done, P will also do  $a$ , which ought not to be done. We do not need (c) to reach this conclusion.

I will now put forward the theories of Bergström, Feldman, Sobel and Tännsjö and sketch how these theories seek to take care of the problems.

For Bergström the solution is to apply (U) only to *relevant* alternative-sets. What then is a relevant set of alternatives? Bergström has proposed several answers to this question (see [1], pp 42-55), but in [3] (pp 124-125) he seems to have decided on the following characterization:

- (R) A is a relevant alternative-set for P in S if, and only if, A is an alternative-set for P in S and, for every A', if A' is another alternative-set for P in S, and A' is an expansion of A, then either (i) A and A' are not (U)-inconsistent, or (ii) there exists some alternative-set for P in S which is an expansion of A' and which is not (U)-inconsistent with A.

An alternative-set A is *an alternative-set for P in S* if the actions in A are performable by P in S. Two or more alternative-sets are *(U)-inconsistent* if the normative conclusions which follow when (U) is applied to these alternative-sets are inconsistent. An action  $a'$  is a *version* (quasi-version) of an action  $a$  if  $a'$  is different from but agent-identical and time-identical (not time-identical) with  $a$ , and it is logically necessary that if  $a'$  is performed, then  $a$  is also performed. If A and A' are different alternative-sets, then A' is an *expansion* of A if for every action  $a$ ,  $a$  is a member of A only if  $a$  or some version or quasi-version of  $a$  is a member of A'.

In Case One it seems that  $\{a, b\}$  and  $\{a', a'', b\}$  cannot both be relevant alternative-sets. If we assume (unrealistically) that these are the only alternative-sets for P at  $t$ , then, since  $\{a, b\}$  is (U)-inconsistent with its expansion  $\{a', a'', b\}$ , and since there is no expansion of  $\{a', a'', b\}$ , the latter must be the relevant alternative-set for P at  $t$ . Thus,  $a$  will be given no normative status by (U), and  $b$  will only be given normative status relative to one alternative-set: it ought not to be done. This resolves (P1) and (P2) and makes clear that (c) does not hold on Bergström's account of alternatives.

Fred Feldman's solution is somewhat similar to Bergström's. Feldman can be seen as restricting the application of (U) to a certain class of alternative-sets, in his case sets of life histories.

Feldman assumes that 'at every moment of moral choice, each agent has available to him a rather large supply of "life histories", each of which is a set of acts exactly coinciding with all he has already done, and the rest of which he is still

able to perform' ([2], p 260) Furthermore, there is said to be a possible world for each possible life history, such that if the possible life history is lived out, this possible world is what will be the case. These possible worlds, 'life history worlds' as it was, can be ordered according to the amount of positive over negative value (utility) they contain. A life history world open for an agent at a specific time is said to be optimific if it's utility is no less than the utility of any other such world. The main thesis of Feldman's 'World Utilitarianism' can now be stated like this:

- (WU) It is right at [a time]  $t$  for [a person]  $S$  to perform  $a$  at [the time associated with  $a$  =]  $t(a)$  if and only if there is a world,  $w$ , such that  $w$  is an optimific life history world for  $S$  at  $t$ , and  $S$  performs  $a$  at  $t(a)$  in  $w$ . ([2] p 266)

In Case One this means that if  $a'$  is performed in the optimific life history world available to  $P$  at  $t$  (we suppose there is only one that is optimific), it is right for  $P$  at  $t$  to perform  $a'$ . Furthermore, since  $a'$  and  $b$  are incompatible, it is not right for  $P$  at  $t$  to perform  $b$ , since  $b$  cannot be performed in the same world as  $a'$ .

Feldman would probably regard  $a'$  as a conjunctive act consisting of  $a$  and  $c$ , where  $c$  is the act of reading the paper on smoking and lung-cancer. And he would hold that if the conjunctive act  $a \cap c$  is done, then of course  $a$  is done and thus it would be right for  $P$  to do  $a$ , provided that  $a'$  is right. Feldman would thus avoid the problems in Case One while accepting a modified form of (c): 'If a conjunctive act is right at a time (performed in an optimific world), then each of it's conjuncts is right at that time too (since they too are performed in this world) '.

But we still do not know what the normative conclusions are in Case One. On Feldman's view, the example is insufficiently precise to determine the normative status of the actions. Nothing direct is said about the utility of the worlds in which the actions are performed. There are two problems: Firstly, Feldman does not say anything about when the consequences of an action are better than the consequences of a given different action. Secondly, if his theory is expanded in a reasonable way with a definition of this relation, the assumptions in Case One are incoherent. The expansion I have in mind is this:

- (e) The consequences of an action  $a$  are better than the consequences of a time-identical and agent-identical action  $b$  iff  $a$  is performed in a world open to the agent of  $a$  at  $t(a)$  with a higher utility than any such world in which  $b$  is performed<sup>6</sup>.

In addition I will say that one action is better than another iff the consequences of the one are better than the consequences of the other.

It now seems impossible to picture how, in Case One,  $b$  could be better than  $a$ , while  $a \cap c$  ( $=a'$ ) is better than  $b$ . For Feldman, the fact that  $a \cap c$  is done in a world implies that  $a$  is done in the same world, and since  $b$  is incompatible with  $a$ , it is *not* done in this world. Suppose that the best world open to  $P$  at  $t$  is  $w$ . If  $b$  is better than  $a$ ,  $b$  is performed in this world. But then  $a \cap c$  is not better than  $b$  since  $a$  is not

---

<sup>6</sup> This principle is faithful to the spirit of utilitarianism in making true the following: a) If one act is worse than an alternative act, it cannot be true that it is right, or ought to be done. b) If an agent must perform either of two actions, and one is better than the other, this action is right, and ought to be done.

performed in  $w$ , nor is  $a \cap c$ , since it cannot be performed if  $a$  isn't. Case One becomes inconsistent.

Jordan Howard Sobel could be seen as applying (U) to the set of lives *securable* at the moment in question. He suggests the following act-utilitarian principle:

- (S) An action  $a$  ought to take place iff (i)  $a$  is contained in a life optimum among lives securable by the agent of  $a$  at this action's first moment (that is, a life optimum among those lives each of which would be secured by some fully specific minimal action open to the agent of  $a$  at the first moment of  $a$ ) and (ii) no agent-identical action incompatible with  $a$  satisfies (i). (See [4], p 196)

This principle calls for some explanation. That a life,  $L$ , would be *secured* by an action,  $a$ , means that if  $a$  is performed,  $L$  is lived out. That an action is *minimal* means that once begun, this action cannot be stopped.

Examples are placing a bet, beginning to swallow a chocolate, some ceases of swallowing a chocolate and beheading by guillotine. An action is *fully specific* if and only if there are no two minimal incompatible actions that are versions of it, in the sense defined on page 3. An action is *contained* in a life if it is true that if the life in question is lived, the action is performed. That a life is optimal means that, in Feldman's words, its life history world is optimific.

It is worthwhile to note that for Sobel, as for Bergström, actions are abstract entities, 'propositions' in Sobel's terminology. Thus, if it is true that  $a''$  is done, it follows logically that  $a$  is done. His theory also implies that if  $a'$  ought to take place,  $a$  ought to take place, whereas  $a''$  and  $b$  ought not to take place. In this respect Sobel's position is similar to Feldman's, and seems to cause the same problems: It is hard to apply (S) to Case One, as none of the actions in Case One seems to be fully specific minimal actions. But, as with Feldman's theory, (S) can be extended with a principle like (e), with the following results: If  $a \cap c$  has better consequences than  $b$ ,  $a$  too has better consequences than  $b$ . If on the other hand  $b$  has better consequences than  $a$ , it also has better consequences than  $a \cap c$ . Thus it cannot be the case that, on Sobel's account,  $a'$  ( $a \cap c$ ) is better than  $b$  which is better than  $a$ . Again, Case One becomes inconsistent.

Feldman's and Sobel's theories challenges Case One. Is there a reasonable interpretation of Case One on which the value-assumptions are consistent? In what sense, for example, is  $a$  better than  $b$ ? I will now try to fill in the contours in Case One with a more detailed story which answers these questions. Here is such a story:

If P has his cigarette and read that article, it will become evident to him that he is risking his health, and he will stop smoking the very same day. (Which would be very good for him.) P's ruining his health will not be evident to him if he smokes without reading that paper, or if he reads that paper without smoking. Thus  $a'$  is better than  $a''$  and  $b$ .

If P has his cigarette, this will make him somewhat drowsy, and he will neither enjoy his favourite TV show as he would do if he did not have this cigarette, nor read anything. This makes  $a$  somewhat worse than  $b$ .

If P does  $a$ , he will in relevant aspects do the same thing as he would if he had done  $a''$ . This means that  $a''$ , as well as  $a$ , is worse than  $b$ .

This story does not seem to be *utterly* unrealistic. But it might be. It might for example be that the law-like conditional ‘If P does have his cigarette (If P does  $a$ ), P will not read that paper’ is weird since P *can* perform  $a'$ , that is, have his cigarette and read that paper. The conditional does not fit – it cannot be true if the assumptions about alternatives in Case One are to be true. To avoid this problem I will assume that (moral) freedom of decision is compatible with determinism. Even if my decision to do something or my doing it is causally determined, it might be a (morally) free decision or act.

Torbjörn Tännsjö holds that act-utilitarianism should be concerned with concrete particular actions, that is, definite processes existing in space-time. It is these very processes that have consequences that can be normatively relevant. If we conceive of actions as concrete in this sense, cases like Case One ceases to be problematic for utilitarianism, Tännsjö holds.

I think Tännsjö would reason as follows about Case One: Suppose that if P does  $a$  (have a cigarette), this will be done during the time-interval  $t(a)$ . Since  $b$  is an alternative to  $a$ ,  $b$  would also be performed during  $t(a)$ , since alternatives are time-identical. And since  $a'$  and  $a''$  are alternatives to  $b$ , they will be performed during  $t(a)$  too. Now, either  $a$  is the same concrete action as  $a'$  or  $a''$ , or it is not. Since  $a$ 's consequences are different from those of  $a'$  and  $a''$ ,  $a$  is not identical with any of these. In that case,  $a'$  and  $a''$  are alternatives to  $a$ , and vice versa. It might be somewhat surprising that P's having a cigarette right now is an alternative to P's having a cigarette and reading that paper right now. But on Tännsjö's view, the concrete action P would perform if P had a cigarette right now could very well be different from the concrete action P would perform if P had a cigarette and read that paper. It might for example be the case that P would not read that paper if P were to perform  $a$ , but P would do this if P performed  $a'$ . And if  $a$  and  $a'$  differ in such a way they cannot be the same *concrete* action.

If  $a$  is an alternative to  $a'$  and  $a''$ , there is only one alternative-set in Case One:  $\{a, a', a'', b\}$ <sup>7</sup>, and in this set, the consequences of  $a'$  are better than those of the other alternatives. Neither (PI) nor (P2) arises. If  $a$  would have been identical with, say,  $a'$ , it would logically speaking have the same consequences as  $a'$ , and the alternative-set in (CI) would have been  $\{a, a'', b\}$  or equivalently  $\{a', a'', b\}$ . (Since ‘ $a'$ ’ and ‘ $a''$ ’ are two names for one and the same concrete action.)

## Case Two

During a period T, P can either  $a$  (have a cigarette) or  $b$  (not have a cigarette). After this, P can during a period T' either  $c$  (read a paper on smoking and lung-cancer) or  $d$  (not read a paper on smoking and lung-cancer). It is also true that P can perform each of the following actions during T +T':  $a \cap c$  (have a cigarette and read that paper),  $a \cap d$  (have a cigarette but not read that paper),  $b \cap c$  (not have a cigarette but read that paper) and  $b \cap d$  (not have a cigarette and not read that paper). These

---

<sup>7</sup>Only this set fulfils the requirements in (AS).

actions are unrealistically assumed to be the only actions P can perform during the rest of his life.

As a matter of fact,  $b$  and  $d$  is done. It is true that  $d$  is done regardless of what P does during T. That is, whether P performs  $a$  or  $b$  makes no difference when it comes to P's doing  $c$  or  $d$ . But it is not the case that P *cannot* perform  $c$  rather than  $d$ , he just doesn't.

These actions' consequences are valued as follows:  $b$  is better than  $a$ ,  $c$  would have been better than  $d$  if  $a$  had been done, but  $d$  is better than  $c$  now that  $b$  has been done. Furthermore  $a \cap c$  is better than  $b \cap d$  which is better than  $b \cap c$  which is better than  $a \cap d$ .

The story behind the value-assumptions is this: Since  $d$  is performed, the only difference in value between  $a$  and  $b$  is that P's health is affected to the negative if he does  $a$ . The pleasure P sometimes gets from smoking is not had in this case – the only cigarette available was a Prince light. Thus  $b$  is better than  $a$ . Next,  $c$  is better than  $d$  if  $a$  has been done, as reading that article after having an unpleasant smoke will make P stop smoking earlier than he otherwise would, and the sooner he stops, the better. If, on the other hand, P did not have that cigarette, he would be practically unaffected by the paper, except for a small uneasiness felt when reading it.

Of the actions done during T +T',  $a \cap c$  is the best. The unpleasant smoke together with the reading will make P stop smoking. This effect will not be had on any of the time-identical alternatives. Second best is  $b \cap d$  as it does not cause any bad health or unpleasantness from smoking, nor any uneasiness from reading. Third is  $b \cap c$  as the discomfort from reading is slightly less bad than the bad health gotten from smoking. The last alternative,  $a \cap d$ , is accordingly the worst.

Bergström would hold that  $a \cap c$  ought to be done during T +T'. If  $a$  had been done,  $c$  would ought to be done, but since  $b$  is done,  $d$  ought to be done. Nothing will be said on the normative status of  $a$  and  $b$ , as neither of them is a member of a relevant alternative-set. The alternative-set  $\{a, b\}$  is not relevant as it is (U)-inconsistent with the set containing the four actions performed during the whole of T +T'. (Bear in mind that there is no expansion of this set.)

Assuming that the actions mentioned in Case Two are the only ones P can perform during his life, Feldman would hold that the life history world where  $a \cap c$  is performed is the best one available at  $t$  (= the beginning of T). It would then be right for P at  $t$  to perform this act, and consequentially it would be right for P at  $t$  to perform  $a$  and to perform  $c$ . As P does  $b$ , it will be right for him at  $t'$  (= the beginning of T') to perform  $d$ . If, however, P had performed  $a$ , it would have been right for him at  $t'$  to perform  $c$ , as  $c$  would be contained in the best life history world available to P at  $t'$ .

Sobel's theory would – under our unrealistic assumption – yield the following results: Actions  $a$  and  $b$  are the only fully specific minimal actions available to P at  $t$ . If P does  $a$ , he will have a bad cigarette, and will gain nothing from it. (P will perform  $d$ , remember.) As P does  $b$ , he will not have these unpleasant effects of smoking. It is then clear that  $b$ , but not  $a$ , is contained in the optimum among lives securable by P at  $t$ . Thus  $b$  ought to be done according to (S). It also follows that  $b \cap d$  ought to be done, as it too is performed in this world. This is very different from the result of applying Bergström's and Feldman's theories to this case, and the reason is that Sobel wishes to take into account mistakes made by the agent at

a time later than the action in question. The reasoning is as follows: If P is to perform  $a \cap c$ , which he ought to do according to theories like Bergström's, he will start off by doing  $a$ , but will then make the mistake to perform  $d$ . He will thus end up doing  $a \cap d$ , which was the worst action performable during T+T'. In Sobel's words,  $a \cap c$  is not securable as it cannot be begun in a way that is not interrupted by a mistake. The action  $b \cap d$  can be begun in such a way and is better than any of its alternatives.

A question naturally leaps to mind when one consider the following two features of Case Two: a)  $a \cap c$  can be performed and b)  $a \cap c$  cannot be begun such that it will be completed. Are these features really consistent? It seems to me that b) implies that  $a \cap c$  *cannot* be performed. To perform an action, one needs to begin the performance of it. If this cannot be done in a way that leads to the action being performed, one cannot perform the action. And this means that the first two paragraphs in the description of Case Two is inconsistent. Questions like these will be discussed in part two and three of this essay.

Tännsjö would hold that if P performs  $b$  or  $a \cap c$  or  $d$ , he has acted rightly. According to Tännsjö, P is here facing a moral dilemma: it is right for him to do  $b$ , but if he does, he will also perform  $b \cap d$ , which would not be right for him to perform. Now, I have a hard time understanding this. It seems that the  $a$  P performs if he does  $a \cap c$  is different from the  $a$  mentioned above, as it is followed by  $c$ . If we assume that  $c$  does not follow upon this  $a$  purely at random,  $c$  should be seen as a consequence of this  $a$ , and a good one too, which make the consequences of (this)  $a$  better than the consequences of  $b$ . Thus it seems that  $b$  isn't right after all. This will be discussed further in part two and three.

I hope that what I have said this far gives the reader an idea of what the problem is about, and what the participants of the debate have said about the problem. In the parts that follow, I will be more critical and seek to settle the issue.

## Criteria of adequacy

In this part of the essay I propose two criteria of adequacy for utilitarian theories. Each of them is a necessary but not sufficient conditions for an adequate theory, I claim. The criteria are:

- (i) The theory should be concerned with performable, that is *effectively decidable* actions only.
- (ii) The theory should count as consequences of actions only what is *determined* by these actions.

I will use these criteria in a discussion of the coherence of Case Two, and in criticizing the theories of Feldman and Sobel. But first some explanation and to some extent a defense of the criteria:

### Effective decidability

In ethics (at least as I conceive of it here), we are concerned with voluntary action, where 'voluntary' is to be understood in a broad sense, covering actions which are the result of subconscious as well as conscious decisions. For P to be able to



perform  $a/c$  it must be possible for him to cause<sup>8</sup> the performance of the action by a decision to perform it. In the given moment of choice, the conditional 'if decision, then action' must be a law-like<sup>9</sup> truth. If P can form a decision to perform an act before performing the act and do this in such a way that the act is performed, I will say that he can *effectively decide* to perform the act in question, and I will furthermore take this to be a necessary and sufficient condition for this act being performable by P.

Someone might be tempted to argue that most actions are not the result of a given decision at a certain point of time. Rather, it is an on-going concern that keeps the action going. This point seems to be mostly terminological. In my vocabulary a decision is a kind of coming to have an intention. For the action decided upon to actually be performed, this intention or decision must be integrated enough, that is 'supported by the person's whole motivational and cognitive orientation'<sup>10</sup> to a certain degree. We can say that the intention must be kept up throughout the act, and that the act is thus kept going by an on-going concern.<sup>11</sup>

A different complication is this: It might be that the performance of an action is only made probable by the decision to perform it. How should we treat such cases? The probabilities we are concerned with here are metaphysical or ontological probabilities, not epistemological. It is not only that because of our limited cognitive or perceptual capabilities we cannot know, given knowledge of the decision and the situation in which it is made, that the action will be performed; there is no true law-like statements to the effect that if a decision exactly like this one is taken in this kind of situations, then an action exactly like this one will be performed. To put it differently: there are at least two possible worlds, exactly coinciding from the beginning of time up to and including the decision, such that in one of them the action is performed while in the other it is not.

The kind of situation I am to analyze then, is not one with which we are used to deal. But we are used to deal with (merely) epistemic probabilities in similar situations. Consider the following case: An athlete longs for a gold-medal. To get it, he will have to train very hard and forsake other things that he finds valuable. But he does not know for sure whether his best efforts will ever get him that gold-medal, since he might get injured, or simply physically unable to run or jump as well as he would need to. It is natural for him to weigh the values of the possible

---

<sup>8</sup> The concept of 'cause' I have in mind here, taken from J. L. Mackie's "Causes and Conditions" (*American Philosophical Quarterly* 2 1965, pp. 245-64) can be spelled out like this: An event X takes place because of an event Y if Y is an insufficient but necessary part of an unnecessary but sufficient condition for X. If X is P's doing  $a$  during T, and Y the decision in the situation of choice at  $t$  ( $=s$ ) to perform  $a$ , an unnecessary but sufficient condition for X might be P's decision in  $s$  to perform  $a$ . P's decision to perform  $a$  is then an insufficient but necessary part of this condition.

<sup>9</sup> It might do with statistical laws (quantum mechanics). But that would be an extension of the common sense conception of choices and voluntary action. See the next section for more on this issue.

<sup>10</sup> Harald Ofstad, *An Inquiry into the Freedom of Decision*, Norwegian Universities Press 1961, p 18.

<sup>11</sup> Sometimes we don't need a very well-integrated intention to start with, even if the action is of a very demanding type. We might during the acting get new motives and knowledge that supports our intention. But the initial intention with it's support was still, in the favoured terminology, a *cause* of the action being performed.

outcomes of trying to get that medal against their probabilities. And I propose that we should do the same when it comes to these 'objective' probabilities. When we consider the normative status of an action and come to know that it is only made probable by the decision, we should not only consider the consequences of our action, but also of possible different outcomes of trying to perform the action<sup>12</sup>, and weigh these outcomes together by taking the value of each possible outcome of the decision in question, multiply with its probability given the decision, and add the products. The sum is the value of the *decision's* consequences or, in other words, the utility of the decision. If it is nonsensical to assign figures to probabilities and values of outcomes and treat them mathematically, the process could be made more intuitively, but along the same lines: intuitive probabilities are weighed against the intuitive values of outcomes, and it is all intuitively added together. As far as I can see, this is how we proceed when we take epistemic probabilities into account, and it seems reasonable to extend this procedure to objective probabilities.<sup>13</sup> (The intuitive method would not give very precise or definite values, but in that case: should we really expect any?) This treatment leaves the action in question without any normative status, which seems intuitively reasonable; to say of an action which is only made probable by any decision to perform it that it is obligatory is to my ears very much like saying of an action that isn't performable that it is obligatory.

Before turning to the next criterion of adequacy, I will consider an argument that might tell against the first criterion of mine. Bergström argues (in [5], pp 97-98) that an action can be obligatory without being performable in this sense. He asks us to consider the following dialogue:

- X 'You ought to get up at six o' clock tomorrow morning.'
- Y 'No, I ought not.'
- X 'Why?'
- Y 'Because I have decided to stay in bed tomorrow.'
- X 'Surely, that is completely irrelevant. It doesn't in the least contradict my statement that you *ought* to get up at six. '
- Y 'Oh yes it does. Since I have decided to stay in bed tomorrow, I shall in fact not get up at six. And even if I were to forget my decision and start to get up, I would immediately remember it again and stop the action short of completion. I would merely pull the blanket over my head and try to get back to sleep.'
- X 'Do you mean to say that you cannot get up at six tomorrow, or that at least it would be very difficult for you to do so?'
- Y 'Not at all. Of course I can get up at six. I do that quite often. But I won't do it tomorrow. I don't want to, and that's why I have decided to stay in bed.'
- X 'Would you agree that there are several different ways in which you could start to get up at six tomorrow?'

---

<sup>12</sup> *Trying to perform the action* will in cases like this always be an effectively decidable action.

<sup>13</sup> One way to understand objective probabilities is to consider them limit cases of epistemic probabilities. Normally, we think it in principle possible to know which of some alternative states will be the case, for instance which side of a coin will turn up when it is tossed in a certain way. Objective probabilities are such that it is in *principle* impossible to know this. In both cases we should act on probabilities.

- Y 'Yes, I suppose so.'
- X 'Would you also agree that there is some way to start which would be followed by the remainder of the action?'
- Y 'Certainly not. There is no way in which I could start to get up which is such that, if I were to start in this way, I would get up at six tomorrow.'
- X 'I see. Well, then I must admit that it is not obligatory for you to get up at six. I was wrong when I said that you ought to do so.'

In Bergström's presentation, the dialogue is meant to show the absurdity in taking securability (in Sobel's sense of the word) to be necessary for obligatoriness. But I think that Bergström would like to use the same kind of argument against my performability-condition. Bergström claims that Y's argument is absurd, and that there is nothing that should motivate X's change of mind.

In the dialogue, Y claims that there is no way to start getting up which will be followed by Y getting up. If this is true, the act of getting up would according to my performability-condition be impossible for Y to perform. This would be a *reductio* of my position since I would have to accept the absurd argument of Y's. Luckily, Y's argument is not absurd in accepting a link between performability and securability, or between performability and effective decidability. The absurdity of the dialogue comes from the weird psychology of its characters or from some very unusual context which is not spelled out. For one, Y's claim that he won't get up at six, no matter what precautions are taken, is surely absurd. He could, for example, presumably act upon the moral advantages of getting up early, or sign a legally binding promise to give away all his money, should he not get up. (Remembering the promise will make him jump right out of his bed quick as ... ) Moreover, Y only considers starting getting up out of habit or due to having temporarily forgotten his decision *not* to get up, but not that he should start getting up as a result of a *decision* to get up. This is an odd oversight. So it seems to me that if Y's claim is to be true, he or the setting of the dialogue must be very strange.

A different but equally striking absurdity is that Y is allowed to argue from his not *wanting* to get up to his not being (in my sense) *able* to get up. In performing a voluntary action, we do what we want to do. If we had acted differently, this would have been because of our wanting to act differently, and it seems that Y in a morally relevant sense could have wanted to get up; had he been slightly less lazy, or just a bit more concerned about other people's well-being, he would have wanted it. Again, either the situation in which X is talking to Y or X himself must be highly unusual. Since this is the case, it is very hard to assess the bearing of our intuitions in this case. If one starts from absurd premises one is likely to reach absurd conclusions no matter how respectable rules of inference or reasoning one is using. I conclude that a fair counter- example to my performability-condition is still wanting.

### **Consequences are *determined* by action**

I believe that an adequate utilitarian theory should count as consequences of an action only those events or states of affairs that are determined by the performance of the action, either logically or causally.<sup>14</sup> (If I eat a lot of food, it follows logically that a lot of food is eaten, and it follows causally that I have had

---

<sup>14</sup>The concept of cause I have in mind here was introduced in footnote 8.

enough to eat.) This is my second requirement which means that things that happen by pure chance (things that are unpredictable in principle) will not be counted as consequences. In [6], Tännsjö does not accept this; at least he argues against Bergström's theory in a way that presupposes that the requirement does not hold.<sup>15</sup> But since he provides no argument against it and as I believe that it is both normatively and descriptively reasonable, I will let it stand.

---

<sup>15</sup>The paragraphs I have in mind is part of an argument for concretism and against Bergström's theory in which alternatives are abstract entities. The argument goes that if we do not accept a concretist account of alternatives, *either* we will have to make a maximally specific description of the action, which is the same as to specify exactly which *concrete* action we have in mind, *or* we will - unacceptably - have to presuppose that (a simple) determinism is true. Tännsjö holds that if we don't make fully specific descriptions of the alternatives or conceive of them as concrete, there will always be versions (in the sense defined on p 3) of any given alternative. And if determinism is false it might as long as there are versions of a given action be some versions of the action that are worse, and some that are better than any alternative to it...

For even if there are no empirical laws to the effect that one version brings about a certain effect if may happen in a concrete unique case that the action *is* followed by this 'effect' (which would not have resulted if some alternative action had been performed instead). ([6], p 6).

This would mean that no action with versions, that is, no actions but the fully specific or concrete ones, could be members of what Bergström calls relevant alternative-sets, and consequently no such action would have normative status. If this is true, only fully specific or concrete actions would have normative status, and this is what Tännsjö argues. Tännsjö seems to hold that even if determinism is false, it might be true that one version of an action is followed by an event which would not have followed if a different version of that action had been performed. I find this somewhat puzzling. So, without making any ontological assumptions, I will use the language of possible worlds to try to clarify things. Tännsjö speaks as if the following could be true:

- (i) In the possible world where the version  $a'$  of the alternative  $a$  is performed, the event  $e$  takes place.
- (ii) In the possible world where  $a''$ , a version of  $a$  incompatible with  $a'$ , is performed,  $e$  does not take place.

However, since only one of  $a'$  and  $a''$  can be performed, at least one of these propositions must be a counterfactual truth, which makes it a law-like truth that if  $a'$  ( $a''$ ) is performed, then  $e$  will (will not) take place. Suppose that  $a'$  is performed in the actual world, then (ii) is a counterfactual truth. But so is (i)! This is because (i) talks about *the* possible world in which  $a'$  is performed. If there is only one such world, it is *eo ipso* a law-like truth that  $e$  follows upon  $a'$ . Tännsjö would probably not assert both (i) and (ii) then. The passage '...it may happen in a concrete unique case that the action is followed by this "effect"' from the quotation above could in this example be taken to mean that  $e$  takes place after  $a'$  is done, while

- (iii) There are some possible world(s) in which  $a'$  is performed and  $e$  does not take place, as well as some where  $a'$  is performed and  $e$  does take place.

In that case it just *happened* that in the actual world, where  $a'$  is done,  $e$  takes place. Consequentially,  $e$  was not *determined* by the performance of  $a'$ . But then it is unreasonable to take the occurrence of  $e$  as a consequence of  $a'$ . The performance of  $a'$  makes it to some degree, say  $p$ , *probable* that  $e$  should take place, but it does not make it true that  $e$  takes

If an action does not determine any outcome, but make several outcomes probable to some degrees, the value of the actions consequences is the sum of the values of the different possible outcomes weighed against the probabilities of the outcomes.

## Criticism

In this part the two criteria from part two will be applied first to Case Two, which turns out to be inconsistent if my concept of performability is assumed, then to the theories proposed by the participants of the debate, of which Feldman's and Sobel's are refuted. The theories of Tännsjö and Bergström are also discussed and put in doubt.

### Case Two revisited

In Case Two, the following is true:

- (i) At  $t$ , P can perform  $a$  and  $b$ , and only these, during T.
- (ii) Regardless of what is done during T, P can at  $t'$  perform  $c$  and  $d$ , and only these, during T'.

Does it follow from (i) and (ii) that

- (iii) at  $t$ , P can perform  $a \cap c$  during T + T'?

The validity of such a deduction seems to be implicit or explicit in parts of the discussion of alternative-sets.<sup>16</sup> But it will be clear that (iii) does not follow from (i) and (ii), neither in general, nor in cases like Case Two. In fact, these three properties together make Case Two inconsistent.

One trivial way in which the deduction from (i) and (ii) to (iii) does not hold *in general* is the following: it might at  $t$  be impossible for P to know that  $c$  will be an option. If so, P can at  $t$  neither decide to perform  $c$ , nor decide to perform  $a \cap c$ , since this involves deciding to perform  $c$ . (Remember that performable acts are effectively decidable.) We could along these lines explain why, in Case Two,

- (iv)  $d$  is done during T', regardless of what is done during T.

On this explanation, P is not very interested in medical research, and when, at  $t'$ , he stumbles upon the up to then unknown article on smoking and lung-cancer, he instinctively puts it aside. But since the explanation implies that  $a \cap c$  is not performable, it would make Case Two inconsistent.

To save Case Two from inconsistency and to show that the deduction holds in it, we need a different but reasonable explanation of (iv) which does not make false one of (i) through (iii). Even though these four features of Case Two are shared by

---

place. That is the work of chance. (Of course, if  $a''$  rather than  $a'$  had been performed,  $e$  would not have taken place, which makes it true that the performance of  $a'$  rather than  $a''$  made the probability of  $e$ 's occurring  $p$  rather than zero.)

<sup>16</sup>For example, Sobel's definition of 'open (=performable) action' (p198 in [4]) makes this deduction valid. The definition is used throughout [4].

several thought-experiments in the debate, no such explanation will be found; consequently many arguments put forward by the participants of the debate rest on shaky ground.<sup>17</sup> (In fact, most of what has been said in [2], [4], [5] and [6] is infected by these incoherent examples.) Here is why Case Two and corresponding examples are incoherent:

If P performs the voluntary action  $a \wedge c$ , he will at  $t$  decide to perform it and, because of this decision, perform it during  $T+T'$ . Suppose that P at  $t$  can and does decide to perform  $a \wedge c$ . He performs  $a$ . but, – alas! – according to (iv) he diverges from his decided path at  $t'$ . Why is that? There can be two kinds of explanations for this: Either the divergence is some sort of indeterministic freak (micro-physical or a God's act of fancy). Or it was predetermined all along.

If the divergence was *predetermined*, it seems that  $a \wedge c$  could not be performed after all: At  $t$ , there were (according to (i)) exactly two ways open for P. One of these,  $b$ , was sure to lead away from the performance of  $a \wedge c$ . The other way was not such that P's decision at  $t$  to perform  $a \wedge c$  caused  $a \wedge c$  to be performed. It will not help to say that 'maybe P didn't try hard enough', or 'his decision was not well-integrated', since the only alternative to this ineffective decision was to go for  $b$ .

If, on the other hand, the divergence was *not predetermined*, the causal link between the decision and the performance is broken. Again, there was no way to cause the performance of  $a \wedge c$  by the decision to perform  $a \wedge c$ .

To sum up: if (i), (ii) and (iv) are true, P cannot effectively decide to perform  $a \wedge c$ , and therefore cannot perform the act.<sup>18</sup> The incoherence of Case Two can show itself in other ways. In the application of Tännsjö's theory to Case Two (see p 10), the performability of  $a \wedge c$  was taken for granted, and feature (i) above was implicitly put in question. Even if  $a$  (having a smoke) was followed by  $d$ , which made  $a$  a poor alternative, there had to be (since  $a \wedge c$  was taken to be performable) a different  $a$  or a different way of having a smoke which would lead to  $c$  being performed, namely the smoking initiated by the effective decision to perform  $a \wedge c$ .

This means that (i) does not hold if (ii), (iii) and (iv) are true. It also means that moral dilemmas of the kind Tännsjö has been discussing in his papers are illusory, based on incoherent thought-experiments. (This does not, of course, detract from the plausibility of Tännsjö's theory.)

### Feldman's theory

Feldman's World-Utilitarianism was formulated like this in his [2]:

(WU) It is right at [a time]  $t$  for [a person] S to perform  $a$  at [the time associated with  $a$  =]  $t(a)$  if and only if there is a world,  $w$ , such that  $w$  is an optimific life history world for S at  $t$ , and S performs  $a$  at  $t(a)$  in  $w$ . ([2] p 266)

In addition, Feldman holds that an act is obligatory for an agent at a time if and only if the act is performed in every optimific life history world open to the agent at the time. Notice that this does not preclude that the obligatory act is performed in life history worlds (open to the person at the time) which are worse than all other

<sup>17</sup>Bergström's Example 3 in [2] is in relevant ways like Case Two, and so is Sobel's Case I ([4], pp 199-200) which also shapes the discussion in Bergström's [5] which in turn plays an important role in Tännsjö's [6] and [7].

<sup>18</sup>Again, a slightly different treatment is needed if P's sticking to the path of  $a \wedge c$  was made metaphysically or ontologically probable by his decision.

life history worlds (open to the person at the time). Nor does it preclude that an action is obligatory even when it is performed in all the disastrous life history worlds (open to...) just because it is performed in the optimistic world, which might be just barely better than a host of other worlds. This means that an action that is obligatory can be a virtual disaster if, as is typically the case, it is performed in 'wrong' world. How can this be implications of a satisfactory theory?

I can think of several answers that Feldman could invoke to his defense, but in the end, none is satisfactory. But let's have a look at an example. Imagine the following: a man is flying with his child in a one-seated aeroplane when he runs out of fuel. He can try to reach a good landing spot by using the altitude and velocity of the plane (ending the flight in a crash resulting in ugly but not permanent wounds) or he can jump (there is a perfectly good parachute in the plane) carrying his child in his arms. He jumps, but since he doesn't release the parachute, he gets paralysed from neck and down. The child dies. If he had used the parachute, it would have landed him and the child safely. Assuming that the optimistic life history worlds open to the man before the jump all included the parachute descent, to jump was at the time right as well as obligatory for him according to (WU). Feldman's first line of defense could then be to claim that the man's performing the action 'to jump' did not determine the fatal landing, and accordingly that the paralysis and the child's death isn't a consequence of the action.<sup>19</sup> The reason should be that the action 'to jump' is too unspecific; it has several versions, some of which contains the parachute descent. Since the natural thing to do if one jumped would be to release the parachute, it might be that for the man to jump determines that the release is very probably performed. If this is the case, it might be right to say that it was right to jump, but that the man screwed up. But suppose instead that if the man jumps, it is very unlikely that he will get to releasing the parachute, or in other words: there are many worlds open to him in which he jumps, but very few in which he does a parachute descent. One would now like to say not that he ought to jump, but that he ought to *parachute* jump or – since it is very uncertain whether he will use the parachute even if he starts jumping with the intention to use it – that he ought to go for an emergency landing.<sup>20</sup> But Feldman's theory would still hold that it is right to jump and even that it is obligatory.

Maybe I have misinterpreted Feldman in giving this kind of an example. There are some indications in [2] to the effect that Feldman's theory is about *concrete* acts; on page 256 he writes that act-utilitarianism is best construed as a theory about concrete acts, understood as 'particular, dated, non-repeatable individuals such as the walk I took yesterday afternoon, the robbery committed last night at Amherst, and my wife's writing a check at noon today'. If Feldman had given some examples of acts that have not been performed, the case would be easier to settle. Even if one talks about particular abstract acts, one can be taken to talk about concrete acts (in Tännsjö's sense) if one mentions only acts that as a matter of fact have been or are performed. Whether Feldman would hold that actions that are not performed are concrete in this sense is still open, but we could ask what difference it would make if Feldman followed Tännsjö on this point.

If Feldman's is a concretist theory, 'the man ought to jump' assigns a certain normative status to a concrete action that exists during a certain time-interval and

---

<sup>19</sup>Assuming that the theory fulfils the second criteria of adequacy.

<sup>20</sup>I think it wrong to say that he ought to parachute jump if the outcome of his trying to parachute jump is this uncertain. See part two, section one (*effective decidability*) above.

within a certain space. Assume first that the concrete action referred to exists from a point of time when the man leaps out of the plane until he reaches the ground, and that the man does not use the parachute. This is the concrete act of jumping which the man in the case performs, and an act like that is presumably not contained in any optimistic life history world open to the man before he jumps, so Feldman does not have to say that what he did was right at the time of the jump. This concretist interpretation of the example seems to have avoided the problem.

But let us instead assume that we refer to a concrete action which exists during a time interval ranging from the moment the man decides that he should jump until his next moment of 'moral choice', presumably shortly after he leaves the plane. During all of this concrete action, the man intends to release the parachute shortly after jumping out of the plane. But, falling through the air he suddenly comes to think it a great or even beautiful thing to be crushed to the ground with the child in his arms (he is mesmerized by a sense of weightlessness, I guess). He *can* pull the release, but he doesn't want to and because of this the child is crushed dead and the man's spine is broken. Still, according to Feldman, the man's jumping could even have been obligatory, since the (concrete) jump could be included in all the best life history worlds open to him at the time, life history worlds where the jump is followed by the parachute release. I find this unreasonable, but let's have a closer look. Assume that the death of the man's child and his own disablement was *caused* by the decision to jump, and furthermore that every alternative concrete act of which it is true that the man jumps with his child in his arms also has as a consequence that they are both crushed. I cannot see how such an act can be right or obligatory. Of course, we have to take into account if the decision to jump caused neither the disaster nor the decision not to use the parachute, but made them probable to some degree. Again, even if the chances for the jump to end in a happy landing are but a tiny fraction of the chances for a disaster, Feldman would tell us that it is right to jump, as the improbable world where the man *does* use the parachute is one of the best life history worlds open to the man. It could at the same time be wrong to do an emergency landing as long as this action isn't contained in the very best world open to the man at the time, even if this action is contained in most of the better worlds, but in none of the worlds where the disaster takes place. To me this seems to be very much against the spirit and the appeal of consequentialism. Feldman could argue that there was nothing wrong with the act of *jumping* – it was the failure to pull the release that was wrong. But then it would be far-fetched to call it an act-consequentialist theory<sup>21</sup>, and act-consequentialist theories are my interest in this essay. I will therefore look for a better theory.<sup>22</sup>

### Sobel's theory

- (S) An action  $a$  ought to take place iff (i)  $a$  is contained in a life optimum among lives securable by the agent of  $a$  at this action's first moment (that is, a life optimum among those lives each of which would be secured by some fully specific minimal action open to the agent of  $a$  at the first

---

<sup>21</sup>Admittedly, Feldman does not call his theory a consequentialist one: utilitarianism is what he is concerned with.

<sup>22</sup>In some *ideal* sense, that is provided that the man did everything else as he ought to, one might say that he acted rightly. Compare Bergström's discussion of ideal notions of 'ought' in [3], especially on page 128.



moment of  $a$ ) and (ii) no agent-identical action incompatible with  $a$  satisfies (i). (See [4], p 196)

I am going to see what (S) does to the example in the last section. But first it is important to understand what it is for an action to be contained in a life. Notice that actions, as Sobel conceive of them, are propositions expressed by sentences such as 'The man jumps' ([4], p 197). An action is now said to be contained in a life L just in case it is true at L's first moment that the action would take place if L were lived;

So actions logically entailed by sets of actions in L are contained in L, and so are actions that, roughly and intuitively, consist of such entailed actions together with parts of what would be their conventional and causal upshots. ([4], p 199)

The picture of a life as a conjunction of propositions with logical implications emerges. If we go by this picture we can generate the following violation of what I believe is the spirit of act-consequentialism: Assume that 'The man jumps out of the aeroplane with his child on his arm, releases the parachute and lands safely on the ground' is an action (call it  $a$ ) contained in the life optimum among lives securable by the man at the first moment of this action. According to Sobel, it ought to take place. But it is entailed by the performance (=making true) of this action<sup>23</sup> that 'The man jumps out of the aeroplane' (call it  $a^*$ ) is performed (made true), wherefore this action ought to be done, too. And here the problem from the last section appears again: Sobel will hold that  $a^*$  ought to be done even if it is contained in some or many lives that are far from optimum: namely the lives of which it is true that the man jumps but chooses not to pull the release. It does not matter whether  $a^*$  is contained in every disastrous life or is most probably performed in a way that ends in disaster, as long as it is contained in this single optimum life.

However, Sobel does avoid one of Feldman's problems as he can treat actions performed after the actions we are considering as *consequences* of actions that are performed now; instead of counting what *could* be done if an action is performed, he considers what *would* be done if the action is performed. This is clearly more in line with my demand that consequences are determined (causally or logically) by the action.

How could Sobel avoid the problem he gets in cases like the one with the jump? The source of the problem is that an action can be contained in several lives, and that only the membership of the best of these lives counts in giving the action a normative status. The most straightforward way to avoid the problem would be to restrict oneself to actions that are contained in only one life, what would be called a *fully specific* action. Another, similar way to avoid the problem is to hold that less specific actions aren't logically implied by more specific actions. The reason could be that the action is individuated not only by what is said to be true about it, but also by what isn't said e.g. that 'The man jumps' is performed does not follow from the fact that 'The man jumps without using the parachute' is performed, since if the man performs the first of these actions, he neither uses the parachute, nor

---

<sup>23</sup>If it is taken as a proposition or set of propositions.

abstains from using the parachute. In other words, the principle of bivalence does not hold for abstract actions on this construction. From this it follows that 'The man jumps' can be an alternative to 'The man jumps without a parachute' (and this seems somewhat contra-intuitive). A different solution would be to somehow weigh together the utilities of the different lives in which the action is contained, presumably by weighing the utility of each life against the probability of that life, given the action. The first two solutions would be somewhat like Tännsjö's: only actions that have no versions are taken into account. The third solution would be a (fairly) new one, I believe.

### **Tännsjö's theory**

Tännsjö does not accept my two criteria, since he would accept the formulation of cases like Case Two. One consequence of this was that a person finding himself in a situation like the one in Case Two faces a moral dilemma: he ought to do  $a \cap c$ , but if he does, he will also perform  $a$ , which ought not to be done. If Tännsjö would come to accept my criteria of adequacy, his theory would avoid this problem, as it does avoid the problems in Case One. Other problems remain, however. In the discussion of Case One I noted that P's having a cigarette would be an alternative to P's having a cigarette and reading the paper on smoking and lung-cancer, since they might be different concrete actions which cannot be performed together. This seemed to be rather contra-intuitive. Consider:

#### *First dialogue*

'I think he should go to the movies.'

'No, he should go to the movies and see "Batman".'

#### *Second dialogue*

'What's bothering you?'

'I do not know what I should do. Either I will go to the movies or I will go to the movies and see "Batman".'

#### *Third dialogue*

'Do you think it a good idea to go to the movies?'

'Yes, if you see "Batman".'

To my mind, and to people I have asked for an opinion, it is clear that the first and the second dialogue are absurd, while the third is quite plausible. But what is the proper diagnosis of the absurdity of the first dialogues?

As I said above, a concrete action like the one the person would perform if he went to the movies can very well be different from the concrete action he would perform if he went to the movies to see 'Batman'. So why is the negative answer in the first dialogue absurd? The least far-fetched answer is that it is very unreasonable to expect the two actions to be incompatible. But this is not the case if the actions we usually have in mind when considering questions like these are concrete. In the second dialogue: why is it strange that someone is trying to make up his mind whether he should go to the movies or go to the movies and see 'Batman'. Presumably since these actions aren't incompatible. And why is the affirmative answer in the third dialogue quite plausible? Again: the actions are not incompatible.

One of Tännsjö's reasons for accepting concretism is that it is the more 'natural' view of actions ([6], p 3). I think the dialogues and what has been said about them makes this claim rather implausible. But Tännsjö's main argument for concretism is that the non-concretist will have a hard time telling us how abstract actions can have consequences. If this proves too difficult, we might have to become concretists in order to avoid the problems, even if concretism isn't the most natural way to conceive of alternatives.<sup>24</sup> My first step in assessing the validity of this argument will be to understand how the *concretist* accounts for the consequences of alternative actions. Tännsjö's idea in [6] seems to be that everything that happens in the world where an action is performed is a consequence of the action (see footnote 17). In part two, section two, we saw that this won't do. Instead I propose the following:

Assume that L is the conjunction of all law-like truths (statistical and/or deterministic). Assume further that the world-description D(c) is the conjunction of everything that would be true of the concrete action c and the world as it is when c is begun.<sup>25</sup> Everything that follows logically from L and D(c) is what follows from the performance of c. Call the conjunction of this R(c). The results of the alternatives c', c'',... c<sup>n</sup> to c<sup>26</sup> are R(c'), R(c''),... R(c<sup>n</sup>). The consequences of c can now be defined as everything which follows from R(c) but doesn't follow from everyone of R(c'), R(c''),... R(c<sup>n</sup>).<sup>27</sup> Now, the same account seems to work for abstract actions. Admittedly, if an abstract action, a, is less specific than the conjunction of all true statements about c, D(a) is less specific than D(c), and consequently R(a) might be less specific than R(c). But I see no *a priori* reason for assuming that consequences of alternatives always should be as specific as the consequences of concrete actions.

Tännsjö would object that hardly any positive state of affairs is necessitated by the performance of an action like 'P's studying philosophy at university U during 1993', even if the action is taken to be performed in a certain environment. Probably only statements like 'P will *not* study psychology at university C' will follow. And this seems to be a *reductio* of the non-concretist position since we usually think that an action like my studying philosophy at Uppsala next year would have more, and more specific, consequences than this. There is a remedy, however. I believe there is an implicit agreement that we consider how the person

---

<sup>24</sup>I am quite sympathetic to concretist accounts of actual and performed actions. But it is a different matter to understand what we are comparing when we think of not as yet performed possible actions or actions that neither have nor will be performed. The later problem seems more interesting when it comes to ethics or problems in the theory of decision.

<sup>25</sup>'the world as it is when the action is begun' signifies a state of affairs exactly similar to the state of affairs in which the actualized of the alternatives is begun in, except from what follows logically and causally from the fact that this different action is begun in it.

<sup>26</sup>the alternatives to c' should here be understood as every action with which c should be compared for it's normative status.

<sup>27</sup>Sometimes philosophers claim that the consequences of an action are those events that would occur if the action is performed, but would not occur if the action isn't performed. This assumes that there is *one* answer to what would happen if the action isn't performed, but since we most often face more than two alternatives, several different things can happen. The definition above solves that problem. A simpler way to solve the problem would be to identify the result of an action (as defined above) with it's consequences. But this would make something that happens in a gambling hall in Hong Kong immediately after my switching of my computer a consequence of my switching, even though I could have made nothing to prevent it.

in question would perform the abstract action, given his capacity, personality and the situation in which the action is performed. We say things like: 'If I would have stayed at my place Friday night, I probably would have watched TV. That's the way it usually happens' or 'If he goes to the movies, he will see "Casablanca". The guy just loves Bogart and Bergman'. If the action we are thinking and talking about isn't actually performed, or if it isn't yet performed, we have to consider the person with a personality somewhat different from the personality he actually had (or has) when the action he actually performed was started. To the very least he would have wanted to do something else, and for some reasons which would not be identical with those that actually made him act. These are complicated issues, but here it suffices to note that there are ways (actually in use, I take it) to expand  $D(a)$  such that there will be something positive, even a lot, to say about the consequences of a particular abstract action. The reductio suggested above is avoided. In fact, when Tännsjö discusses how we are to refer to possible concrete actions, he suggests that we refer to them by using expressions like 'what he would do if he were to decide to (perform the abstract action to) go to the movies'. (Compare [6], p 9) If this would suffice to point out a specific *concrete* action, it will also allow us to get a sufficiently – perhaps fully – specific world-description  $D$ ('He goes to the movies'), and in turn sufficiently specific consequences  $C$ ('He goes to the movies'). If, on the other hand, it does not suffice to point out concrete actions, a reasonable account of how one can refer to a concrete action is lacking.

### Bergström's theory

Consider the following alternatives: either I go to the movies ( $a$ ), or I don't ( $\neg a$ ). The consequences of  $a$  are better than those of  $\neg a$ . What reasons could we have for being discontented with this alternative-set when we are inquiring into what I ought to do? I suggest that the following two reasons are important when it comes to expanding alternative-sets:

i) We believe that there are ways to go to the movies that have better consequences than simply going to the movies. If I went to the movies I might for example come to see 'Batman'. But it happens that 'The third man' would give me more pleasure and intellectual stimulation; 'I go to the movies and see "The third man"' ( $a'$ ) has better consequences than  $a$ . In the following dialogue, an expansion of  $\{a, \neg a\}$  is made:

#### *Fourth Dialogue*

'Do you think I should go to the movies?'

'Yes, if you see "The third man".'

The new alternative-set is  $\{a', a'' \neg a\}$ , where  $a'$  is the act of going to the movies and see 'The third man' and  $a''$  is the act of going to the movies without seeing 'The third man';

ii) We believe there are *versions* of not going to the movies that are better than going to the movies. If I decided not to go to the movies, I would end up on the sofa eating cheese doodles and watching 'Miami Vice'. But I could for example do some useful studying ( $\neg a'$ ), which would have better consequences than just not going to the movies.

#### *Fifth Dialogue*

'I think I should go to the movies rather than not.'

‘No, you should not. It would be better if you did some studying tonight.  
[And, for reasons under ii):]  
‘I still think I should go to the movies. They give “The third man” at the  
Odeon tonight, and I know it's damn good.’

This story could go on for a very long time. The person I am talking to might claim that if I stay at my place and study Nietzsches ‘The will to power’, this will be even better than watching Orson Wells in that great film of his, and I could retort that I could read some Nietzsche on my way to the movies...

Considerations like those under ii) led Bergström to applying the utilitarian principle only to relevant alternative-sets (as defined by (R)). In [1] (p 52) he proposed another necessary condition for relevant alternative-sets roughly corresponding to the first reason for expanding alternative-sets. It has however been shown by Włodzimierz Rabinowicz<sup>28</sup> that Bergström's theory implies that in some cases two alternative actions are obligatory. Since the proof is somewhat complicated I will not give any details of it, but the main idea is that by starting from different alternative-sets we can prove the obligatoriness of two incompatible actions. As far as I can see, the proof is a *reductio ad absurdum* of Bergström's theory.

There is also a different problem with Bergström's theory. It might for example have better consequences to perform the abstract act of going to the movies than to perform the abstract act of going to the movies and see “Batman”. Sometimes it might be hard to change one's mind once a decision is made even though new information should make a version not decided upon more attractive and it is generally a good thing to leave room for spontaneous reactions to the environment. But none of Bergström's proposed criteria for relevant alternative-sets could give one a reason to ‘step back’ from one alternative-set to a set with less specific actions. Because of this, and because of the problem discovered by Rabinowicz, we have to find a better theory.

## Suggestions

Remember the question in the beginning of the essay: what are *the* alternatives in a situation of deliberation and planning? This question could be interpreted in the following two ways:

- (1') What actions should be compared with regard to their consequences when the act-utilitarian criterion of rightness is applied?
- (1'') What alternative-set is it rational to chose from in a situation of deliberation and planning?

I am well aware of there being several other, perhaps interesting, interpretations of the question, but I will concentrate on these.

To answer the first question, we need an act-utilitarian criterion of rightness. I propose the following:

---

<sup>28</sup>‘Utilitarianism and conflicting obligations’ (*Theoria* 1978)

- (UR) An action,  $a$ , is right if and only if (i) no alternative to  $a$  has better consequences than  $a$ , (ii) no version of  $a$  has better consequences than  $a$ , and (iii)  $a$  is not a version of an action with better consequences than  $a$ .

An action  $b$  is an *alternative* to an action  $a$  if and only if  $a$  and  $b$  are (i) performable in the same situation, (ii) agent-identical and (iii) incompatible. 'Version' means the same as 'version or quasi-version' in Bergström's sense. Notice that the two first conditions in (UR) correspond to the two reasons for expanding alternative-sets that were discussed in the section on Bergström's theory. The third condition implies that it might be right not to choose a more specific action; sometimes it is better to decide *how* to perform the action after having chosen to perform it or after having begun the performance of it. Not leaving any room to account for this was one of the drawbacks of Bergström's theory.

I believe that (UR) answers question (1'): an action should be compared with all its alternatives, all its versions and all actions of which it is a version. The criterion of rightness could be supplied with the following:

- (UO) An action,  $a$ , ought to be performed/is obligatory if and only if (i)  $a$  is right and (ii) there is no alternative to  $a$  which is right.  
 (UW) An action,  $a$ , is wrong if and only if (i)  $a$  is not right, (ii) there is no version of  $a$  that is right.

The second clause in (UW) is meant to do justice to our intuitions vis-à-vis the fourth dialogue in the section on Bergström's theory. Let us assume that my going to the movies has worse consequences than my staying home, but that my going to the movies to see 'The third man' has better consequences than both these actions. Let us furthermore assume that there are no other possible alternatives to, or versions of, these actions. It is now reasonable to hold that I ought to go to the movies to see 'The third man'. But it is not reasonable to hold that it is wrong to go to the movies, since I ought to go to the movies and see 'The third man'. Neither is it reasonable to hold that it is right to go to the movies, because of this action's consequences: if I go to the movies (period), I will in fact see 'Batman', and this will have worse consequences than going to the movies to see 'The third man'. That is why the person in the dialogue answers 'Yes', but adds 'if you see "The third man"' which in effect is to say that I ought to go to the movies to see 'The third man'. He doesn't answer my question since the action I ask him about is *too unspecific* to have any normative status.

Suppose that we find ourselves in a situation of deliberation or planning, and that we want to act like good act-utilitarians (if there is such a way to act). We are confronted with some alternatives. Is it rational to choose the best one of *these* alternatives, or should we look at some other set, for example an expansion of this set? With the reservation that the meaning of 'rational' I will use here might be different from the one Bergström uses, I suggest that we should take into account versions of our original alternatives, or actions of which our original actions are version if we think such an action has better consequences than any of our original actions. This might be a good answer to (1").

Another way to see the problem is this: our taking into consideration new actions is itself an action. We then ought to do this if and only if this action satisfies the criteria in (UO). It is right to do it if and only if our changing the

alternative set is right according to (UR). It is wrong to do it if it is wrong according to (UW).

The first of the two questions which started this essay has been answered. The second question was how a coherent formulation of act-utilitarianism should look like. I believe that my three criteria of rightness, obligatoriness and wrongness respectively form the basis of such a formulation, given other ideas that I have put forth, especially ideas concerning performability and consequences.

### References

- [1] L. Bergström, *The alternatives and Consequences of Actions*, Stockholm 1966.
- [2] F. Feldman, "World Utilitarianism" i K. Lehrer (ed) *Analysis and Metaphysics*, Dordrecht, Holland 1975.
- [3] L. Bergström, "On the Formulation and Application of Utilitarianism", *Noûs*, 10 (1976).
- [4] J. H. Sobel, "Utilitarianism and Past and Future Mistakes", *Noûs*, 10 (1976).
- [5] L. Bergström, "Utilitarianism and Future Mistakes", *Theoria*, 43 (1977).
- [6] T. Tännsjö, "The morality of Abstract Entities", *Theoria*, 44 (1978)
- [7] T. Tännsjö, "Moral Conflict and Moral Realism", *The Journal of Philosophy*, 82 (1985).