Consequentialism and Rational Choice: Lessons from the Allais paradox

Bruno Verbeek, Leiden University

§1 A test for consequentialism

A normative theory of choice is consequentialist if it has the following two components. First, it contains a principle that ranks states of affairs from best to worst. Secondly, it requires the agent to produce the highest-ranking state of affairs she is in a position to produce. That is, an action is required if and only if it will produce the best possible outcome in the circumstances. Many consequentialists will be inclined to add that an action is required if and only if it will produce the best expected outcomes. In this essay, I restrict myself to the latter definition of consequentialism. Consequentialism, thus understood, is a well-recognized feature of many moral theories. Classical act utilitarianism is consequentialist, but so are various forms of egalitarianism and ethical egoism. What is not so often realized is that consequentialism is at the heart of normative theories of rational choice, such as game theory and expected utility theory. These theories of rational choice differ from the other consequentialist theories in their standard for ranking states of affairs. However, they share with all consequentialist theories the commitment to the maximization of expected value.

Defenders of consequentialism in moral philosophy often maintain that this commitment to maximization is neutral about value. Regardless of what one might identify as the correct standard to order states of affairs from best to worst, once this standard has been stated it is possible to maximize. Similarly in decision theory: once the agent has ranked the available options according to her preferences, she is expected to
maximize. It seems then that consequentialism puts no constraints whatsoever on what goes into the determination of the ranking from best to worst. If this claim is correct, it is a very strong argument in favor of consequentialism in moral theory and rational choice theory. Therefore, one can think of this claim as a test for consequentialism. If it is indeed true that consequentialism is neutral as the defenders maintain, then the burden of proof is on the opponents of consequentialism. They will have to show that, in spite of its openness to values, it is a bad idea.

In this essay, I will argue that not all versions of consequentialism pass this test. The commitment to maximization of expected value, as it is practiced in expected utility theory and other consequentialist normative theories of choice, precludes the identification of certain values. More precisely, I will argue that rankings of states of affairs in terms of regret or disappointment cannot be accommodated in expected utility theory and other consequentialist (moral) theories. Therefore, we should either rule out regret and disappointment from rational and moral consideration or give up certain versions of consequentialism and with it expected utility theory. This latter alternative does not mean that we should not consider outcomes in rational and moral decision making altogether. However, it does imply that we cannot always characterize the consideration of outcomes as maximizing the expected value.

§2 Consequentialism and maximization

Consequentialism, as it is defined, requires the maximization of expected value. To do this, the agent should be able to determine the value of the prospects of her possible actions. A prospect is a lottery over the possible outcomes of an action with their respective probabilities. Note that a prospect can consist of just one outcome on this definition. This is the case when one is sure of the result of one’s action. The standard
way in which the value of prospects is determined is through the summation of the values of each possible outcome of an action weighted with its respective probability. Let us call this method the expectation method. Decision theory teaches us that this method requires that the ranking from the best to the worst outcome conforms to the axioms of expected utility theory. In particular, this ranking should satisfy ordering, independence and continuity. Ordering ensures that a ranking of all the possible outcomes from best to worst can be made. However, this is insufficient to guarantee that one can use the expectation method for determining the expected value of an action. To do that, we need to be able to assume that the value of the prospect of an action can be ranked according to the individual outcomes it contains. Independence requires just that. Suppose \( x, y \) and \( z \) are prospects. Independence says that \( x \) is at least as valuable as \( y \) if and only if the lottery \([x, p; z, 1-p]\) is at least as valuable as the lottery \([y, p; z, 1-p]\). (Where \([x, p; z, 1-p]\) stands for the lottery that gives \( x \) with probability \( p \) and \( z \) with probability \( 1-p \) and \( 0 \leq p \leq 1 \).) If, in addition to these ordering and independence, the ranking of prospects satisfies continuity one can determine the value of each prospect by the method described above. What is more, it also provides the tools to determine the value of individual outcomes. That is, we can represent the value of both outcomes and prospects on a numerical interval scale. Such a scale cannot determine absolute value but it can determine how valuable an outcome or prospect is in relation to the most valuable or least valuable outcome. This is all that is necessary to use the expectation method.

The expectation method is traditionally associated with the representation of preferences: the domain of expected utility theory. The classic representation theorems all roughly claim the following: if we can assume that the agent’s preferences satisfy ordering, independence and continuity, then we can represent her preferences with the expectation method. Several authors have gone further. They claim that expected utility theory has normative status: it specifies the requirements of rationality. An agent has
rational preferences to the extent that they satisfy the axioms of expected utility theory. If this claim is correct, a rational agent is committed to the expectation method.

However, the expectation method is not necessarily limited to the domain of (rational) preferences. It is an attractive method for all consequentialists who are committed to the maximization of expected value. For example, Derek Parfit and Frank Jackson think of it as the best method for consequentialists to determine what is right in circumstances of uncertainty.\(^9\) John Broome has defended the use of the method and clarified its exact status both within theory of rational preferences and within theory of the good. The reason these consequentialists are attracted to the method is not just the relative simplicity of the expectation method. There are deeper grounds as to why a consequentialist might adopt the method. John Broome has argued in several places the axioms of the expectation method describe the ‘structure of goodness’.\(^{10}\) They characterize the logical features of the ‘\(_A\) is at least as good as \(_B\)’ relation. If this is correct, consequentialists, just like rational agents, are committed to the expectation method.

§3 Consequentialism and dynamic choice

The expectation method requires that the ranking of best to worst satisfies the independence axiom. The acceptability of this method, therefore, depends critically upon (among other things) the acceptability of the independence axiom. In 1953, Maurice Allais published a lengthy essay criticizing the plausibility of this axiom.\(^{11}\) This sparked a debate that continues until today about the normative status of the axioms of expected utility theory.\(^{12}\)

The debate took a crucial turn when Peter Hammond showed that the most criticized axioms of expected utility theory, ordering and independence, can be deduced from consistent choices in dynamic decision problems, i.e., those situations where the agent
has to make a number of choices before reaching the desired outcome. Following Howard Raiffa, Hammond represents such choices as decision trees: schematic representations of the series of choices leading to the outcomes that are open to the agent. Hammond’s basic idea is that one can piece together how to rank the outcomes and prospects by determining which plans and which plan continuations are acceptable in such decision trees. Hammond shows that if the agent is committed to what he calls ‘consequentialism’, the ranking of outcomes will satisfy both ordering and independence.

Hammond’s result is very important. It appears to give the axioms of standard expected utility theory a more secure grounding than had been thought possible. Ordering and independence, according to Hammond, are theorems of a consistent application of ‘consequentialism’ in decision trees, rather than logically independent assumptions about the ranking of outcomes.

However it is important to realize exactly what Hammond meant by ‘consequentialism’. The connection between the standard understanding of consequentialism as the maximization of expected value and Hammond’s sense of ‘consequentialism’ is opaque at best. Edward McClennen has argued that Hammond’s rather vague sense of ‘consequentialism’ should be understood as the conjunction of four distinct principles of choice over time. These principles of choice over time are sufficient to deduce the two fundamental axioms of expected utility theory: ordering and independence.

In what follows, I will refer to McClennen’s principles as principles of dynamic consequentialism. In the remainder of this section as well as the next, I will give an exposition of the principles of dynamic consequentialism and their relation to the independence axiom. This will set up the framework for my argument against the neutrality of consequentialism. The exposition is somewhat technical and may seem to lead us far from the problem that I announced at the beginning of this paper. However, I
beg the reader to bear with me. In the end I will show why this is relevant to determine the neutrality of those versions of consequentialism that are committed to the expectation method.

An example of a decision tree will be useful to understand the scope and force of the principles of dynamic consequentialism. The trees in Figure 1 are examples of decision trees. A tree is a collection of nodes, which can be subdivided in decision nodes (the squares in Figure 1), chance nodes (the circles in Figure 1) and terminal nodes, which are connected by branches. The decision nodes represent a point in time where a choice must be made. The circles are those points in time where uncertainty about chance events is resolved. The branches represent the sequences of choice and chance events before reaching an outcome \( (x, y \text{ or } z) \) at the terminal node of the tree. A plan is a specification of the agent’s moves at all choice nodes that the agent can reach, given the relevant chance events, by making the earlier moves determined by the plan.\(^{16}\) The agent is supposed to formulate her plan at the beginning of each tree. That is why each tree starts with a decision node \( n_0 \).\(^{17}\) In Figure 1a there are two possible plans. Plan 1 is ‘if ‘up’ happens, choose ‘up’ at \( n_2 \)’, whereas plan 2 is ‘if ‘up’ happens, choose ‘down’ at \( n_2 \). Each plan leads to an outcome, depending on the conditioning chance events. The set of outcomes that could be reached by the execution of a plan are the associated outcomes of that plan. For example, plan 1 in Figure 1a has the associated outcomes \( \{x, z\} \). If this set of associated outcomes is not a singleton, the resulting outcome of the implementation of the plan will depend on the intervening chance events. The prospect of a plan is a lottery that has the elements of the set of associated outcomes of that plan as prizes with their respective probabilities. Thus, plan 1 in Figure 1a has the prospect \( [x, p; z, 1-p] \), whereas plan 2 has the prospect \( [y, p; z, 1-p] \).
The principles of dynamic consequentialism characterize the relation between the acceptability of a plan and the acceptability of its prospect under ideal circumstances. These circumstances are, first, that there is no change of information during the execution of a plan. Secondly, there are no constraints on the deliberational abilities and capacities of the agent. That is, the agent has sufficient time and all the necessary computational abilities to make complex decisions. Throughout this paper, I shall assume that these conditions hold.

There are four principles of dynamic consequentialism. The first condition is that of simple reductionism (SR). It is best illustrated with Figure 1b and 1c. In those situations the agent just has to make one choice before all uncertainty, if any, is resolved. All the agent’s plans, therefore, consist of just one ‘up front’ choice. Simple reductionism says that the acceptability plan in such trees is reducible to the acceptability of its related prospect. More precisely:

\[(1) \text{Simple reductionism (SR): Suppose that in a tree } T \text{ the set of associated plans } S \text{ is such that for each plan } s \in S \text{ only one single ‘up front’ choice is needed for its implementation. In such trees } s \text{ is acceptable if and only if its related prospect is at least as good as all other feasible prospects.}\]

Clearly, simple reductionism should be acceptable to all consequentialists, regardless of whether they are committed to the expectation method or not. It is nothing but a
reformulation of the consequentialist maxim to maximize expected value in situations where the agent has to make just one initial choice.

However, such situations are exceptional at best. More often than not, an agent needs to make a sequence of choices before reaching an outcome and often uncertainty is only partially resolved during that sequence. The second condition says that this ought not to make any difference. That is, in any tree with identical prospects, the acceptable plan(s) in each tree will lead to the same prospects(s). More precisely:

(2) Irrelevance of the resolution of uncertainty (IRU): let T be any decision tree. Let T* be a tree where the agent just has to make one ‘up front’ choice. Suppose that T and T* have exactly the same prospects. Then any plan s ∈ S is acceptable in T if and only if its prospect is also acceptable in T*.

In other words, it should not make any difference if an agent decides in a context where she has to make a sequence of choices or if she decides in a context where she makes a single choice up front. As long as the set of prospects is the same, the acceptable plan(s) in both cases should have the same prospect(s). IRU requires a parallel between the selection of plans in Figure 1a and Figure 1b. The choice of plans should be such that the acceptable plan in both trees leads to the same prospect. Just as SR, IRU should be acceptable to any consequentialist. Failing to treat 1a and 1b equally implies that considerations other than the (value of the) outcomes determine one’s choices. SR and IRU together express the idea that a consequentialist should judge plans by their prospects and by nothing else.

It might be objected that the timing of the resolution of uncertainty is not irrelevant. For example, one might prefer to find out whether it actually rains before deciding to go on a picnic or not. However, this misses the point of IRU. IRU is a principle to assess the acceptability of plans. A plan is defined as the ex ante specification of the agent’s moves at all choice nodes that the agent can reach, given the relevant
chance events, by making the earlier moves determined by the plan. In the case of the decision whether or not to go on a picnic, the agent might plan to go if it does not rain and stay at home if it rains. However, the prospect of that plan (i.e., [be at home if it rains; have a picnic if the sun shines]) is typically not available if the uncertainty about the weather conditions can only be resolved after the agent’s decision to go on a picnic or not. For in that context the agent only has the prospects [have a picnic in the rain; have a picnic in the sun] and prospects [stay at home if it rains; have a picnic if it rains]. IRU only applies to trees with identical prospects.

The next condition specifies the relation between an acceptable plan and its components. In a decision tree, the agent goes through a series of nodes. A plan is a specification of what choice is to be made at each future choice node given the chance events. A plan is composed of plan continuations: feasible sequences of choices from the agent’s present choice point in the tree. Dynamic consistency states that a plan is acceptable if and only if all its continuations are acceptable.

(3) Dynamic consistency (DC): for any tree T and any plan \( s \in S \) in T, \( s \) is an acceptable plan in T, if and only if for any point \( n_i \) that can be reached should the agent implement \( s \), the plan continuation of \( s \) in \( n_i \) is an acceptable plan continuation.

This means that the acceptability of a plan should not change as the agent moves through the tree. The series of choices that the agent makes should be consistent with the plan that she adopts at the outset of the decision problem. Note that this condition is plausible only if there is no change in information about the shape of the tree, the value of the outcomes or the intervening chance events. In those cases, it might very well be required to change your mind and form a different plan. However, under the assumed ideal conditions there are no such changes.
This condition is plausible independent of whether one is consequentialist or not. It is implicit in the notion of a plan. A plan comes with a commitment to execute it, other things being equal. If you know in advance you will not carry out your plan, why bother planning? This condition will be acceptable to most people, consequentialists or not.  

The final condition specifies which plan components, understood as plan continuations, are acceptable. Dynamic consequentialism regards each plan continuation as if it is a completely new choice problem. A plan continuation is acceptable, if in the corresponding de novo decision tree, the corresponding plan is acceptable. This is the requirement of separability (SEP). Let T be a tree and T(n_i) be a separate tree, identical to T from n_i onward. Let s(n_i) stand for the continuation of the plan s in T from n_i on. Every plan continuation available in T from n_i on is an available plan in T(n_i). Then:

(4) Separability (SEP): s(n_i) is an acceptable plan continuation if and only if its corresponding plan in T(n_i) is also acceptable.

Therefore, in each subsequent choice node of a tree the agent is to choose as if she were to face the selection of plan de novo. So if n_2 can be reached through the implementation of an acceptable plan in the tree of Figure 1a, the acceptable plan continuation from n_2 on should be identical to the acceptable plan in the tree in Figure 1c. SEP captures part of the forward-looking perspective of consequentialism.  

As we shall see, separability is a deceptive principle. Apart from its formal ramifications, there are at least two ways in which it could be interpreted. On the one hand, separability could mean that one should forget what happened before. In that case, one treats one’s choice at n_2 in Figure 1a as absolutely identical to that in Figure 1c. That is, one ‘strips’ what happened before n_2 as well as the shape of the tree until n_2 from
one’s beliefs. In that case one faces the choice at \( n_2 \) really \textit{de novo}.\footnote{23} On the other hand, one can think of separability as regulating what one should consider in choosing. Then it is not the case that the agent is supposed to forget; rather she should \textit{disregard} what happened before. The agent has access to all the information about the shape of the tree as well as how she got to \( n_2 \), but is supposes to disregard that in deciding how to choose from \( n_2 \) on.\footnote{24} This is a subtle difference that will be relevant when we come to discuss the neutrality of separability in section 8.

Summarizing, we have four conditions that specify a standard of consistency in the rational acceptance of plans. Simple reductionism demands that the acceptability of a plan in a tree with just one ‘up front’ choice node corresponds to the acceptability of its outcomes. IRU requires that the acceptability of a plan correspond to the acceptability of its related plan in a tree with just one ‘up front’ choice node. Dynamic consistency requires that the acceptability of a plan correspond to the acceptability of its components, its continuations, in each choice-node. Separability, finally, requires that the acceptability of plan continuations correspond to the acceptability in related \textit{de novo} trees.

\section*{§4 Independence and dynamic consequentialism}

These four principles are jointly sufficient to derive the central axioms of standard expected utility theory. If there exists a plan in a finite decision tree that is acceptable on these principles, the ranking of the outcomes must be such that it satisfies ordering as well as independence. Moreover, if the ranking of outcomes satisfies ordering and independence and the acceptable plan satisfies SR, IRU and DC this plan also satisfies SEP. Therefore, separability is necessary. Since I will be particularly concerned with the status of the independence axiom for consequentialist theories of choice, I will give an informal proof as to how acceptance of the principles of dynamic consequentialism
commits the agent to accepting independence. A similar proof of the derivation of ordering can be found in the appendix.25

Remember that independence is satisfied if it is the case that for any three prospects \(x, y\) and \(z\) it is the case that \(x\) is at least as good as \(y\) if and only if \([x; p; z, 1-p]\) is at least as good as \([y; p; z, 1-p]\). To determine whether the principles of dynamic consequentialism commit the agent to rank prospects in accordance with independence, we need to consider two cases. First, the case that \(x\) is at least as good as \(y\). Secondly, the case that \([x; p; z, 1-p]\) is at least as good as \([y; p; z, 1-p]\).

(1) Suppose that \(x\) is at least as good as \(y\). SR tells us that, therefore, in Figure 1c ‘up’ is an acceptable plan. Turning our attention to Figure 1a, we see that there are two possible plans, with two related prospects. On the one hand plan 1: ‘if ‘up’ at \(n_1\) happens, then ‘up’ at \(n_2\)’, and on the other hand plan 2: ‘if ‘up’ at \(n_1\) happens, then ‘down’ at \(n_2\’.

SEP tells us that at \(n_2\) is ‘... up’ an acceptable plan continuation, because in the de novo situation (Figure 1c) ‘up’ is an acceptable plan. Since an acceptable plan consists of acceptable plan continuations, DC tells us that plan 1 is an acceptable plan in Figure 1a, since it calls for ‘..up’ at \(n_i\).

The two available plans in Figure 1a have two related prospects, namely plan 1: \([x; p; z, 1-p]\) and for plan 2: \([y; p; z, 1-p]\). The tree of Figure 1a corresponds to the tree of Figure 1b. In both trees there are just two feasible plans and the same related prospects. IRU tells us that if plan 1 is an acceptable plan in Figure 1a, the plan with the same prospect in Figure 1b must acceptable as well. That is, since \([x; p; z, 1-p]\) is the prospect of an acceptable plan in Figure 1a, the plan (‘up’) in Figure 1b must be acceptable since it leads to \([x; p; z, 1-p]\) as well. This acceptable plan must satisfy SR, since in Figure 1b there is just one up front choice. This means that \([x; p; z, 1-p]\) must be at least as good as \([y; p; z, 1-p]\). Therefore, if \(x\) is at least as good as \(y\) and the acceptable plans in the trees of
Figure 1 satisfy the four principles of dynamic consequentialism, it follows that \([x, p; z, 1-p]\) must be at least as good as \([y, p; z, 1-p]\).

(2) Suppose that \([x, p; z, 1-p]\) is at least as good as \([y, p; z, 1-p]\). We can run a similar argument. This ranking means that in Figure 1b (‘up’) is acceptable (SR). Then, by IRU, plan 1 must be acceptable in Figure 1a. DC tells us that the plan continuation ‘... ‘up’ at \(n_2\)’ must be acceptable in Figure 1a. Then, by SEP, ‘up’ must be acceptable in the *de novo* situation (Figure 1c). SR tells us that, therefore, \(x\) is at least as good as \(y\). (1) and (2) together demonstrate that the principles of dynamic consequentialism require that the ranking of prospects satisfies independence.

Note that similar arguments can be formulated if \(y\) is at least as good as \(x\) and \([y, p; z, 1-p]\) is at least as good as \([x, p; z, 1-p]\). Moreover, similar arguments can be made independent of in which tree we start. Once we know at least one acceptable plan in any of them, the repeated application of the principles of dynamic consequentialism will bring us to the conclusion that the ranking of the outcomes satisfies independence.

A crucial assumption in the proof above is that there are at least three trees with identical outcomes. This seems an innocent assumption because I did not specify \(x\), \(y\) or \(z\). However, should it turn out that the shape of the tree affects the outcomes it may become less innocent. For then, it might turn out that there is –say – only one tree that has *these* outcomes. If that is the case, the principles of dynamic consequentialism cannot bring us to the conclusion that independence is required.26

I have argued that SR, IRU and DC should be acceptable to all consequentialists. We have seen that independence is necessary for the use of the expectation method. What about SEP? As it turns out, SEP is a necessary condition for any consequentialist committed to the expectation method. That is, if the agent plans in accordance with SR, IRU and DC and her ranking from best to worst satisfies independence, her plans will have to satisfy SEP. The proof is relatively simple. Suppose that \(x\) is at least as good as \(y\)
and – as is required by independence – \([x, p; \xi 1-p]\) is at least as good as \([y, p; \xi 1-p]\). Then SR tells us that the plan ‘up’ in Figure 1b and 1c is acceptable. Since the prospects of all available plans in Figure 1b are identical to those in 1a, IRU tells us that the plan ‘if ‘up’ happens at \(n_1\), go ‘up’ at \(n_2\)’ is acceptable. DC requires that a plan consists of acceptable plan continuations. Therefore, it must be the case that ‘up’ at \(n_2\) is an acceptable plan continuation in Figure 1a. Since the continuation of the tree in Figure 1a is identical to the complete tree of Figure 1c, SEP is satisfied if and only if going ‘up’ is an acceptable plan in Figure 1c. Since \(x\) is at least as good as \(y\), SR tells us that this is the case, QED.\[27\]

We can conclude that there is an intimate connection between independence and dynamic consequentialism. Since expected utility theory is committed to independence, it is committed to these four principles. Similarly for consequentialism: since the preferred method of aggregation (i.e., the expectation method) can only work if the ranking of prospects of actions satisfies independence, consequentialists are committed to the four principles of dynamic consequentialism. At the beginning of this essay I announced that I would show that expected utility theory as well as any form of consequentialism that is committed to the expectation method is not neutral about what is of value. I can now make the announcement a bit more exact. Since there is an intimate connection between the four principles of dynamic consequentialism and independence, we should investigate if these principles are neutral about what is of value.

§5 A violation of separability

In order to determine whether the principles of dynamic consequentialism are plausible requirements of rational and moral choice, we should look at typical violations of independence in the light of the four requirements of acceptable planning. Perhaps the most famous of all such violations is the ‘paradox’ first described by Maurice Allais.\[28\]
Suppose an agent faces the choice between two prospects A and B. A is a lottery [\$5M, .10; \$1M .89; \$0 .01] and B is \$1M for sure (M stands for million):

Suppose that an agent argues to herself that one million is better than nothing at all. Of course, five million would even be better, but the chance of that happening is not so high in lottery A. What is more, if she chooses A there is a (very) small chance that she ends up with nothing at all, while she could have assured herself of one million. Therefore, this agent judges B better for her than A. Suppose the same agent is the confronted with the choice between C and D. C is a lottery [\$5M, .10; \$0 .9] and D is a lottery [\$1M, .11; \$0, .89]. She believes that C is better than D, since C gives her a virtually equal chance of five million as opposed to one million.

This seems to be a perfectly plausible set of judgments. However, agents with Allais-type preferences violate the independence axiom. Remember that independence requires that preferences over compound lotteries should be determined only by those elements that differ between the two lotteries. We can represent the preferences over the two pairs of lotteries as the table in Figure 2.

<table>
<thead>
<tr>
<th>States of affairs</th>
<th>p (prob. = .01)</th>
<th>q (prob. = .10)</th>
<th>r (prob. = .89)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$0M</td>
<td>$5M</td>
<td>$1M</td>
</tr>
<tr>
<td>B</td>
<td>$1M</td>
<td>$1M</td>
<td>$1M</td>
</tr>
<tr>
<td>C</td>
<td>$0M</td>
<td>$5M</td>
<td>$0</td>
</tr>
<tr>
<td>D</td>
<td>$1M</td>
<td>$1M</td>
<td>$0</td>
</tr>
</tbody>
</table>

Looking at the table in Figure 2, independence says that we should disregard what happens under r in each of the two choices since the outcomes under r are the same. If B is better than A for the agent, the prospect [\$1M, .11] (which occurs under p and q) must
be better than \([0, .01; 5M, .10]\). However, when considering the choice between C and D the agent judges the former better than the latter. So an agent who prefers B to A and C to D violates independence.

As follows from the claims from the previous section, the agent with Allais preferences violates the principles of dynamic consequentialism as well. To be precise, the plans such an agent follows do not satisfy either SEP or IRU. We can illustrate these claims by presenting the choice over the lotteries A and B, and C and D respectively, as two decision trees.

![Figure 3a](image1)

![Figure 3b](image2)

How does this agent violate SEP? In Figure 3a there are two available plans to the agent. Plan A ‘if ‘up’ happens at \(n_1\), then ‘up’ at \(n_2\)’ and plan B ‘if ‘up’ happens at \(n_1\), then ‘down’ at \(n_2\). These plans have two related prospects that are identical to lotteries A and B described above. Since the agents prefers lottery B to A, IRU and SR imply that plan B is acceptable and plan A is not. DC stipulates that a plan consists of acceptable plan continuations. So if B is acceptable, this implies that ‘down’ at \(n_2\) in Figure 3a is the uniquely acceptable plan continuation. In Figure 3b, on the other hand, the two available plans are C (if ‘up’ happens at \(n_1\), then ‘up’ at \(n_2\)) and D (if ‘up’ happens at \(n_1\), then ‘down’ at \(n_2\), which have lotteries C and D as their related prospect. The agent’s preferences of C over D together with the requirements of SR and IRU, imply that plan C is uniquely acceptable. That means, by DC, that the plan continuation ‘up’ at \(n_2\) is uniquely acceptable.
Remember that SEP says that a plan continuation is acceptable if and only if it is an acceptable plan in the related de novo tree. Now the de novo trees in Figure 3a and 3b from $n_2$ onwards are identical. SEP rules out that the plan continuation (‘up’ at $n_2$) is uniquely acceptable in Figure 3a while at the same time the continuation (‘down’ at $n_2$) is uniquely acceptable in 3b. Therefore, agents with Allais-type of evaluations of prospects violate SEP.

Alternatively, it could be argued that the agent violates IRU. First, note that Figure 2 is in fact the situation where the agent has to make one ‘up front’ choice after which all uncertainty is resolved. By hypothesis the agent judges B better than A but C better than D in Figure 2. Above I assumed that this pattern would be repeated in the trees of Figure 3. However, there is room for arguing that this assumption is unwarranted. From the point of $n_0$ in Figure 3a and Figure 3b, the agent needs to plan on what to do if she were to reach $n_2$. At $n_2$, the agent faces the choice between $1$M for sure and [[$5M^{10/11}$; $0^{1/11}$] in both trees. In Figure 3a and 3b, the agent has to plan for the same contingency. How could there be a difference in the acceptable plan in those trees? In other words, whereas it might be reasonable to judge B better than A, but C better than D in the one ‘up front’ choice case, the decision as to which plan is acceptable in the trees in Figure 3 will not follow that judgment. The implication of this argument is that the timing of the uncertainty of $n_1$ is relevant from the point of view of $n_0$. If it follows the choice that is to be made at $n_2$, then there will be a difference between Figure 3a and 3b but not if it precedes this choice. Hence, it is not separability, but the indifference of resolution of uncertainty that is at stake. Note that this line of argument assumes separability. The claim that since at $n_2$ both trees are identical therefore there ought to be no difference in the acceptable plan continuation in both trees, is tantamount to separability. In conclusion then we have two alternative diagnoses of the Allais paradox in its dynamic form. Either such an agent treats the choice in Figure 3 as equivalent as that of Figure 2 (i.e., assumes
IRU) and ends up violating separability, or the agent regards the acceptable plan continuations in the trees Figure 3 as identical thus violating IRU.

In section 3, I argued that IRU is a central assumption for any form of consequentialism. Rejecting IRU amounts to claiming that considerations other than the value of the outcomes can determine the rational and moral choice. Whereas non-consequentialists might be willing to accept this implication, consequentialists should balk at the very idea. Therefore, I will not consider whether consequentialists should give up IRU rather than separability. Whereas there are good reasons for consequentialists to be concerned about separability, rejecting IRU is anathema.

§6 Individuating the outcomes

There may be room for avoiding these conclusions about the status of separability. Many philosophers, social psychologists and rational choice theorists would argue that the choices of the agent in Figure 3a and 3b are consistent with the principles of dynamic consequentialism. They claim that neither separability nor IRU are violated. For, so the disputants argue, there is no structural resemblance between plan A in Figure 3a and plan C in Figure 3b. In plan A the agent faces the choice between accepting the lottery and regretting her choice or being disappointed when she gets $0M on the one hand and receiving $1M on the other. However, in plan C, the agent does not face this possibility of regret or disappointment. Chances are that she will get $0M anyway, so there is no reason to regret her choice or be disappointed with an unfavorable outcome. In other words, the description of the constituting outcomes of each prospect is incomplete. If we can find more detailed individuations of the outcomes, it will turn out that the Allais preferences do not violate independence or separability. When we adjust these, as is done in Figures 4a and 4b, we see that the de novo tree from $n_2$ onwards in each figure leads to a
different set of outcomes. Going ‘up’ at $n_2$ in Figure 4a and ‘down’ at $n_2$ in Figure 4b will not violate separability. Similarly for IRU: since the acceptable plan continuations in both trees need not be identical, there is no violation of IRU if the agent displays the same pattern of choices in the case of one ‘up front’ single choice node tree. Choosing B in Figure 4a and C in Figure 4b is, therefore, compatible with dynamic consequentialism.

![Figure 4a](image1.png) ![Figure 4b](image2.png)

Moreover, when we correct Figure 3 as this response suggests, we get a different table that does not imply a violation of independence (see Figure 5). Making this point in his discussion of the Allais-paradox, John Broome argues:

All the rationalisations of Allais’ preferences point out a good or bad feeling you may experience. This feeling ... can rationally be taken into account, as well as the money prizes, in determining your preferences. Let us write this feeling into our table of lotteries.

<table>
<thead>
<tr>
<th>States of affairs</th>
<th>$p$ (prob. = .01)</th>
<th>$q$ (prob. = .10)</th>
<th>$r$ (prob. = .89)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$0M$ and regret/disappointment</td>
<td>$5M$</td>
<td>$1M$</td>
</tr>
<tr>
<td>B</td>
<td>$1M$</td>
<td>$1M$</td>
<td>$1M$</td>
</tr>
<tr>
<td>C</td>
<td>$0M$</td>
<td>$5M$</td>
<td>$0$</td>
</tr>
<tr>
<td>D</td>
<td>$1M$</td>
<td>$1M$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

Figure 5
What Broome and other authors ask us to do is to rethink the individuation of the outcomes in the face of apparent violations of the axioms of expected utility theory and dynamic consequentialism. In doing so, they invoke the very same claim that consequentialists in moral philosophy make: consequentialism is neutral about the good.

§7 Trivializing consequentialism: an axiological solution?

There are reasons to be worried about the force of this third response. The most common concern is that this move robs both the theory of rational choice and consequentialism in general of its normative content. Suppose we can find individuations of the outcomes that explain away every apparent violation of separability, IRU or the independence axiom. It seems that this would save these principles as axioms of rational and moral choice. However, we will be left with principles that cannot be violated. Similarly with consequentialism in general, understood as the injunction to maximize. Under some individuation of the outcomes any act can be interpreted as a maximizing one. What good is a requirement of a normative theory of choice if it does not constrain choice? Separability, IRU and independence become empty constraints because every choice will satisfy it under a suitable individuation of the relevant outcomes.

There has to be a principled way to prevent this from happening. Typically this is done by restricting the individuation of outcomes. The most obvious place to look for such restrictions is in the underlying theory of value. A good example of this strategy is that of John Broome. In Weighing Goods he argues that we need a theory that tells us how we should individuate the relevant outcomes of our choices in order to avoid the emptiness of consequentialism.\textsuperscript{31} Such a theory would help us in determining how far we are allowed to go with re-describing the relevant outcomes. Broome proposes the following principle for this individuation:
**Principle of individuation by justifiers:** outcomes should be distinguished as different if and only if they differ in a way that makes it rational to have a preference between them.  

If we apply Broome’s principle to the case of the Allais paradox, we get the following result. If regret is a justifier, this agent does not violate consequentialism. If regret is not a justifier, this agent has irrational value judgments and, therefore, violates consequentialism. This way we have a principled restriction on what can count as the correct individuation of the object of choice.

There are several problems with Broome’s response. First, Broome’s principle does not work in all decision trees. For it is possible that one and the same outcome is part of two different prospects. For example, the outcome $z$ in Figure 1a is part of two prospects: $[x, p; z, 1-p]$ and $[y, p; z, 1-p]$. Let $y$ and $z$ both equal 0 and let $x$ equal $10$. One can imagine that $z$ in the former prospect (receiving $0$ while one could have had $10$) is disappointing, whereas it is not disappointing in the latter prospect, for then one would have had nothing otherwise. Suppose that disappointment is a justifier in Broome’s sense. Then we have conflicting individuations of the very same outcome $z$. It is both disappointing and not, depending on the prospect of which it is part.

A second problem with Broome’s principle is that it compels the consequentialist, both in rational choice theory and in moral theory, to a form of realism in the value theory if the principle is to do the work it is supposed to do. Broome is the first to acknowledge this. Purely subjective theories of value, such as a preference satisfaction theory of well-being that gives the agent authority over what is of value, are ruled out by this move. Here is an illustration of this claim. Suppose that one holds that a prospect A is better than a prospect B for an agent if and only if the agent prefers A to B. Suppose that our agent dis-prefers regret. Then she will treat the prospect [$0, .01; 5M, .10$] as a
different prospect than \([0 \text{ and regret, .01; } 5M, .10]\). According to Broome’s principle she is only allowed to do this if it is rational to prefer the former to the latter prospect. Whether or not this is a rational preference is determined by the theory of value. Only if the former is better than the latter by the lights of that theory, is the agent allowed to make such a distinction. If we use the above-mentioned simple value theory we get the result that it is rational to prefer the former to the latter if and only if the agent prefers the former to the latter. This leads us right back into the problem that Broome’s principle is supposed to solve. Only if the value theory does not give the agent exclusive authority as to what is of value will the principle do its work. I am not claiming that this is wrong. All I want to signal is that this rules out those consequentialists who have a subjective theory of value.\(^3\)\(^4\) In addition, it is sure to alienate many rational choice consequentialists. In that field subjectivism of the sort I described is standard.

The third worry is that there is a scope problem with Broome’s principle. It is not clear whether Broome believes that there is just one correct individuation of an outcome in a particular situation for all agents, or that he believes that the correct individuation can vary between agents.\(^3\)\(^5\) The former seems too strict to be plausible. Surely, what makes a rational, justifiable difference between outcomes for me depends to some extent on the things that I care about, and this can vary between agents. However, if we allow variation between agents we run into another problem. For now it looks as if it is what the agent cares about that determines what counts as a relevant outcome. Then we get the conclusion that if the agent prefers to avoid regret, regret is \textit{ipso facto} a rational justifier. This leads us right back into the normative emptiness that Broome’s principle was supposed to solve.

There may be a way around this difficulty. Suppose that Broome were to claim that the tokens of value differ between agents, thought they are all of the same type. For example, suppose our principle of justifiers specifies that frustration is a bad thing. Then
any frustrated preference is an intrinsic bad. There would be divergence between agents as to what is frustrating since there is a divergence in what they prefer, but it would all be instances of frustration. If this is plausible, we can stick with the strict reading of the principle of Broome’s principle. However, I am not convinced that this is a good move. It may get us out of the scope problem, but brings us back to the original problem for which Broome’s principle was supposed to be the solution. For now it looks as if we can incorporate divergence of justifiers between agents, by re-describing the justifiers in a sufficiently general way. That is, the very same problem of the normative emptiness of consequentialism in rational and moral choice is repeated in the value theory.

Let me make myself clear. I am not saying that Broome’s principle should be rejected. Perhaps one can find plausible answers to all three problems. More work needs to be done. However, axiological solutions, like the one Broome proposes, clearly assume that there is a strict separation between the axiological principles constraining the individuation of outcomes and the formal principles of moral and rational choice. The assumption is that these latter principles do not put any constraints on what could possibly count as a rational justifier. Once again, we run into the assumption that consequentialism is neutral about value. It is time we assess the plausibility of this claim.

§8 Separability, regret and disappointment

At this point, we need to make some distinctions to identify the exact nature of the bad feeling that is to ‘save’ dynamic consequentialism in the Allais paradox. I proposed two candidates for this feeling: regret and disappointment. Exactly what does it mean to say that the agent regrets receiving $0 in Figure 3a and not in 3b? Regret is the (bad) feeling that results from considering what would have been the outcome, given what has happened, had one chosen something different. To regret one’s choices means that one
wishes one had chosen otherwise given what has happened. Regret should be distinguished from *disappointment*. Disappointment is the (bad) feeling that results from comparing what would have been the outcome, given what one has chosen, had something else happened.\(^3^7\) Disappointment, therefore, means that one wishes some other state of affair would have come about, *given what one has chosen*. Another way of expressing this distinction is to say that one regrets one’s choices but is disappointed with the way things turn out. Notice that both regret and disappointment are *ex post* judgments. They arise when one has reached the terminal node in a tree. Looking from that point at the entire tree, one experiences regret or disappointment. So in order to anticipate them when deciding on one’s plan, one should at \(n_0\) imagine oneself to be at each final node and ask what one would feel if one was there. If we employ these notions of regret and disappointment, we do not get an individuation of the outcome of \(0\) that would ‘save’ separability or IRU. In Figures 3a as well as 3b, receiving \(0\) is both regrettable and disappointing.\(^3^8\) \(0\) is regrettable, because given that at \(n_1\) ‘up’ happened and at \(n_3\) ‘down’ happened, one would have been better of if one had chosen to go ‘down at \(n_2\); one would have ended up with \(1M\) rather than nothing. Note that this is true in both trees. \(0\) is also disappointing, because if at \(n_3\) ‘up’ had happened, one would have had \(5M\) rather than nothing.

Suppose, however, that there is an individuation in terms of regret or disappointment that would ‘save’ separability. Suppose that receiving \(0\) in Figure 4a attracts a bad feeling whereas receiving \(0\) in Figure 4b does not. It is clear that the strategy of individuating the outcomes can only work if separability is neutral about the particular individuations that would ‘save’ it. I do not believe that separability is neutral in this sense.

If this bad feeling is going to make a difference between receiving \(0\) in Figure 4a and receiving \(0\) in Figure 4b, it must be that the branch from \(n_1\) leading to \(1M\) in
Figure 4a makes the difference. Let’s call this bad feeling *partial regret* or *partial disappointment* because it focuses only on chance events preceding the choice node, unlike normal regret and disappointment which focuses on all chance events preceding the terminal outcome. The idea is that receiving $0 in Figure 3a is partially regrettable because, given that ‘up’ happened at n₁, the choice for ‘down’ at n₂ would have been a better choice since it does not expose the agent to risk. In Figure 3b, $0 does not attract this partial regret. Given that that ‘up’ happened at n₁, the choice for ‘up’ at n₂ is not particularly risky, since most of the risk has already been borne. Similarly with partial disappointment: given the choice for a plan that requires to go ‘up’ at n₂, receiving $0 is disappointing in Figure 3a, for if ‘down’ happened at n₁, the agent would have had $1M. In Figure 3b, however, there is no room for partial disappointment, since the agent would have had $0 anyway if ‘down’ had happened at n₁.

As we have seen, separability requires that the acceptability of a plan continuation depends upon the acceptability of the related plan in the *de novo* choice situation. The presence or absence of branches in the tree now no longer accessible ought to be of no concern for the determination of the plan continuation. In other words, in planning whether to go ‘up’ or ‘down’ at n₂, the different branches at n₁ ought to be disregarded. That means, that the very thing that causes the agent to anticipate partial regret or disappointment is to be disregarded as she settles on her plan. The very basis for the judgment that $0 in Figure 4a is partially regrettable or disappointing, is to be disregarded. Depending on one’s exact interpretation of separability this implies one of two things.

First, suppose one takes separability to mean that one ought to ‘forget’ at n₂ how one got there. All information about the shape of the tree, the plan one has settled upon and what could have been the case had certain choices and chance events turned out differently is to be ignored. Under this interpretation, separability entails that any
ascription of partial regret or disappointment to $0 is *unintelligible* because the informational basis for this ascription is removed from consideration.

Suppose on the other hand that one takes separability to mean that though one has access to information about the shape of the tree as well as how one got at $n_2$, this information is irrelevant for the determination of the correct plan continuation. Under this interpretation partial regret may be intelligible but should not make a rational or moral difference: ‘$0 and partial regret’ is to be treated as indifferent to ‘$0’. Therefore, separability functions here as a constraint on rational justifiers. It requires the agent to be indifferent about two outcomes. Note however, that this constraint is not based on any axiological commitment. Rather, it is the direct implication of separability.

This shows that separability is not neutral about individuations that implicitly refer to the branch ‘down’ from $n_1$. Whether this individuation is stated in terms (partial) regret, disappointment, or another feeling, separability cannot accommodate them. Given the relation between separability and independence, this conclusion carries over to expected utility theory as well as other forms of consequentialism that use the expectation method. All these forms of consequentialism are not neutral with regards to what is of value. There is at least one consideration, partial regret or disappointment, which is incompatible with the expectation method. The implications of this result go beyond Allais’ example. It implies that any characterization of outcomes that depends on the presence of alternative branches earlier down the line cannot be admitted into the considerations of this kind of consequentialism.

§10 The value of regret and disappointment

However, this is not enough to put the claimed neutrality of separability under pressure. Consequentialists could argue that separability is neutral with respect to what is of value.
It remains an open question whether context dependent considerations such as regret, disappointment and partial regret are valuable. If these feelings are not plausible candidates for being valuable, the incompatibility I demonstrated above is beside the point. So what is it about these feelings that make them proper objects of moral and rational avoidance?

In one sense it is obvious that disappointment and regret are not valuable. So far, I have been arguing that otherwise rational and reasonable agents will avoid experiencing such feelings. Consequentialism holds the view that values should be promoted. Avoidance hardly counts as promotion. Therefore, if consequentialists should be concerned about these feelings, it must be that they stand in some relationship to what is truly valuable. This relationship can be of two kinds. First, it could be that these feelings are intrinsically bad. That is, they are a negative correlate of what is intrinsically valuable. For example, suppose one subscribes to a hedonist value theory and claims that the only thing that is valuable is pleasure. Disappointment and regret could be intrinsically bad in that they directly diminish one’s pleasure.

I am not convinced that this is the best way to defend the claim that regret and disappointment are the proper object of avoidance. For example, several authors have argued that the good is essentially fragile. One can fail to attain it, through no (rational) fault of one’s own. For example, the love for one’s children is a good thing, but it is bound to turn into hurt and pain if the child comes to die prematurely. Attempts to eradicate this fragility, for example by not having any children, are counter-productive. They preclude the very thing that allows for this good. A less high-minded example is the value of winning a lottery. Part of the value is the thrill of winning. Such a thrill would be absent if it were a sure thing that one would win. The potential for disappointment or regret is exactly what makes this thrill possible. If we apply this reasoning to the Allais paradox, we would have the conclusion that the potential disappointment or regret of
$0M does not diminish the value of prospect of choosing A in Figures 2 and 3. Instead it is an essential part of it. Therefore, there are reasons to doubt that regret and disappointment are intrinsically bad.

Alternatively, it could be that regret and disappointment and other such feelings do not themselves diminish what is of value, rather, they track a bad that diminishes the value of the prospect. This is in line with modern cognitive conceptions of emotions according to which emotions are responses that have a world-to-mind fit, much like normal beliefs. For example, fear is an appropriate response to danger and the reason why one should avoid fear is that one should avoid danger. Of course, just like a belief may be false, an emotion may be misplaced. Emotions do not follow values perfectly, but where they do it is reasonable to consider them in one’s plans and decisions.

If we apply this to the emotions of regret and disappointment in the Allais paradox we need to ask what bad do these emotions trace here? One answer in this context could be that these feelings trace risk. Risk, so one might argue, is the negative correlate of security. Psychological research suggests strongly that considerations of security are decisive for the choice for the $1M in the first pair of choices in the Allais paradox. However, this still does not fit the phenomenology of these emotions completely. Regret and disappointment are resultant emotions; they are not prospective. So if regret and disappointment trace risk or the absence of security, why are they not present at \( n_{0} \), when the choices need to be made? Therefore, it is probably best to think of them as a resultant by-product of another emotion, which is prospective in this sense. Hope is a plausible candidate. This is a forward-looking emotion. In deciding to buy a lottery ticket, one hopes for a favorable outcome. If this does not happen it is possible that one is disappointed or regrets one’s choice. On this account hope traces security. Failure of reaching this, the result of risk, can turn hope in to regret or disappointment. I
believe that these reflections are sufficient to show that these feelings have a place in a plausible theory of value.

§11 Are consequentialists committed to separability?

So far, I have argued that the use of the expectation method excludes considerations of regret and disappointment, or rather, risk and security, because of its commitment to separability. This spells trouble for expected utility theory as well for those consequentialists in moral theory who, like John Broome and Derek Parfit, believe that the expectation method is correct. The argument shows that it is not an appropriate method to determine the goodness of the prospects of one’s actions. According to the independence axiom, the value of a prospect is a function of its individual outcomes. Each outcome contributes to or distracts from the value of the prospect independent of the other outcomes. The Allais paradox illustrates why this is sometimes wrong. In determining the value of the prospect of plan A, the regret or disappointment of receiving $0 is the result of it being part of a prospect in which the agent could have received $1M or even $5M. That means that the disvalue of $0 to the prospect as a whole is not determined by that outcome alone. It is co-determined by the presence of other individual outcomes in the prospect. Separability, and with it independence, requires us to ignore these effects. I conclude that expected utility theory and the expectation method in moral theory should be rejected.

We should pose a more general question here. Is there any reason to suppose that consequentialists should subscribe to separability? Recently, a number of authors have put forward arguments that seem to deny this. Nicholas Sturgeon has made a distinction between two kinds of forward-looking. Consequentialism can be forward-looking in the sense that the past and what could have been the case is irrelevant in deciding what to
do. Let’s call this the strong sense. On the other hand, consequentialism can be forward-looking in the sense that it is concerned with future costs and benefits. That is, consequentialism ignores sunk costs. Let’s call this the weak sense. Sturgeon claims that consequentialists are committed to the weak, but not to the strong sense of forward looking. He rejects the strong sense on grounds that it would preclude just about any value that requires information about the past for its identification. Sturgeon is thinking of examples the value of saving money for one’s child to go to college. If one is to ignore the past, one cannot identify this child as one’s own child and, consequently, cannot recognize the value that is at stake.

The strong sense of forward looking is close to the implication of separability that I have been criticizing. However, I am not sure that it is possible to make a sharp distinction between this implication of separability and the injunction to ignore sunk costs. The reason is that sunk costs sometimes contribute to the (dis-)value of an outcome. Here is an example. Ned McClennen once told me the story of two conference participants who had the bad luck to be last in line at the conference dinner buffet. They already paid the $20 fee for this dinner. As time went by, the line did not seem to move fast enough to have dinner and make it to the evening session. One of the two suggested to go out and get a hamburger from a nearby fast food franchise. It would probably cost them $5 and it would save them the hassle of standing in line and they would make it in time for the next session. The other responded: “I am not going to pay $25 for a hamburger!” Clearly, to the second participant the sunk cost of having paid the fee for the conference dinner mattered in the assessment of the value of the hamburger.

The same problem is there when we think of the decision as a decision tree. In Figure 3, the ascription of regret or disappointment is impossible if one ignores sunk costs. The fact that I could have received $1M, if at $n_1$ ‘down’ had happened, is a sunk
cost (or rather, a sunk benefit) that informs the judgment that receiving $0 is disappointing or regrettable. Therefore, Sturgeon’s distinction is not mutually exclusive.

We can reformulate the distinction in a way that captures Sturgeon’s intentions somewhat better. There are two ways in which a particular outcome can be characterized with a reference to what happened in the past. On the one hand, an outcome can be characterized in terms of the causal process that led to it. This is causal backward looking. For example, ‘my son gets a treat’ is an outcome that refers among other things to the particular history that he and I share. On the other hand, an outcome can be characterized in terms of what could have been the case. This is counterfactual backward looking. For example, the outcome ‘my son is disappointed’ refers, given the analysis of disappointment in section 8, to what alternative outcomes my past choices could have brought about. Similarly, ‘I do not regret having a child’ refers to the alternative outcomes I could have realized, given how the world turned out to be. The crucial distinction is that the causal backward looking does not make any reference to counterfactual states of affairs (whether they would result from past chance events (disappointment) or past choices (regret)), whereas counterfactual backward looking does. That is, to understand that my son is my son it is not necessary to wonder what would have happened if he turned out to be a girl. The mere fact that I stand in this special causal relation to him, is sufficient to fully characterize the outcome and identify the appropriate values. Causal forward looking, the denial of causal backward looking, is the real target of Sturgeon’s criticism of the strong sense of forward looking. Counterfactual forward looking, the denial of counterfactual backward looking, covers all cases of weak forward looking, including those cases where the strong and weak senses overlap. I suspect that Sturgeon claim really is that a consequentialist is committed to counterfactual forward looking but not to causal forward looking. We could then reformulate separability so as to avoid causal forward looking, restricting it to
counterfactual forward looking. However, that would still exclude counterfactually backward looking values like regret and disappointment. In other words, if we are to understand consequentialism as the commitment to counterfactual forward looking, as Sturgeon seems to claim, the argument of section 8 still stands.

There are, however, consequentialists who go further. For example, Philip Pettit has argued that consequentialism is not committed to subsequentialism, the view that only future costs and benefits matter.44 I take Pettit to be claiming that consequentialism is not committed any form of forward looking. If this is how consequentialism should be understood, consequentialists are not committed to any form of separability. As a result, independence and the expectation method are not on their list of commitments either.

Such an understanding of consequentialism is immune from the type of criticism that I raised against it. However, it comes at a cost. By rejecting separability in decision making over time, consequentialists such as Pettit not only give up independence, they also give up ordering. In particular, they allow for the violation of transitivity. That is, \( x \) may be better than \( y \) and \( y \) better than \( z \), but \( z \) better than \( x \). The reason is quite simple. Imagine a tree that has three outcomes, \( x \), \( y \) and \( z \) at its terminal nodes, in which the agent first is required to choose for \( y \) or a further choice between \( x \) and \( z \). Suppose that in this tree the plan leading to \( x \) is uniquely acceptable. Since separability is rejected, this has no implications for how to choose in a simple tree with just \( x \) and \( z \) as its final outcomes.45 Non-separable consequentialism, just like separable consequentialism, will need a method to assess the expected value of the prospects of actions. However, if there is no reason to expect that the ranking over the outcomes will be transitive, it is simply impossible to identify the best outcome: what ever outcome one picks, there will always be a better one. Thus, the rejection of separability seems to lead us to the conclusion that the correct consequentialist response to values need not be maximizing. This, I submit, is a major embarrassment to consequentialists.
There are ways out of this awkward conclusion. Suitably refined formulations of the conditions of acceptable planning may help us avoid this conclusion.\textsuperscript{46} Moreover, economists and decision theorists have formulated alternatives to the expectation method. Some of these do not require independence; others allow for violations of transitivity.\textsuperscript{47} It goes beyond the aims of this essay to discuss these alternatives. However, it should be clear that the consequentialist claim that the rightness of an action is completely determined by the expected outcomes it produces is far from unproblematic when we look at decision making over time. Similarly, the claim that consequentialism is strictly neutral with respect to value is doubtful. If anything, this is one of the major lessons from the Allais paradox.
Appendix: dynamic consequentialism entails ordering

The principles of dynamic consequentialism are sufficient to guarantee that the ranking over the domain of outcomes satisfies ordering. Let $O = \{x, y, \ldots, z\}$ be the set of outcomes and $S$ and $S^*$ be proper subsets of $O$. Let $R$ stand for the ranking relation ‘is at least as good as’. $R$ satisfies ordering if and only if:

1. All outcomes are connected (or comparable) under $R$: $\forall x, y \in O: xRy \vee yRx$;
2. $R$ satisfies transitivity: $\forall x, y, z \in O: (xRy \land yRz) \rightarrow xRz$;
3. $R$ is consistent under contraction and expansion: suppose $S = \{x, y\}$, then $\forall x, y \in O: (xRy \text{ in } S \subseteq O) \leftrightarrow (xRy \text{ in any } S^* \subseteq O)$, such that $S \subseteq S^*$. (That is, if $x$ is at least as good as $y$, this judgment is independent of what else is an available outcome of choice.)

I will infer each of these requirements with the aid of the trees in Figure 7.
1 Comparability

Consider the decision tree in Figure 7a above. Suppose that in Figure 7a the plan ‘up, up’ is acceptable. Then, by DC, at n1a the plan continuation ‘..., up’ is acceptable. From SEP we can conclude that in Figure 7a’ the plan ‘up’ is acceptable. SR allows us to conclude that faced with the two outcomes x and y, x is at least as good as y.

Since we assumed that ‘up, up’ in Figure 7a is acceptable, by IRU, ‘up, up’ in Figure 7c must be acceptable. (Figures 7a and 7c have the same normal form representation, therefore IRU allows us to conclude that in their extensive form representation, the acceptable plans in Figure 7a must have the same prospects as the acceptable plans in Figure 7c.) This implies (by DC) that at n1c in 7c ‘..., up’ is an acceptable plan continuation. SEP tells us that this can only be the case if ‘up’ is an acceptable plan in Figure 7c’. From SR we can conclude that, faced with the choice between x and z, x is at least as good as z. So x is connected to all elements in the outcome space. If there is an acceptable plan in Figure 7b’, y and z are connected as well for the same reason. Since this is assumed all along (see the discussion in section 3 and footnote 28), we can conclude that the ranking over the outcomes x, y and z satisfies comparability.

The proof above proceeds from the assumption that ‘up, up’ is an acceptable plan in Figure 7a. However, it does not matter whether ‘up, up’ is acceptable in Figure 7a or another plan. As long as there exists at least one acceptable plan in Figure 7a, the
repeated application of the principles of dynamic consequentialism to all trees ensures that the ranking of all the outcomes is connected.

2. Transitivity

The principles of dynamic consequentialism ensure that the ranking is transitive. Suppose that \( x \) is at least as good as \( y \), and \( y \) at least as good as \( z \). Then, if transitivity holds, it must be the case that \( x \) is at least as good as \( z \). If \( x \) is at least as good as \( y \), the plan ‘up’ must be an acceptable plan in Figure 7a’. This means that in Figure 7a either in \( n_1 \) the plan continuation ‘..., up’ is part of an acceptable plan, or \( n_1 \) is not on the path of an acceptable plan (by DC and IRU).

I will now prove that the \( n_1 \) is part of an acceptable plan in Figure 7a. Suppose, for reductio, that this is not the case and that in Figure 7a the uniquely acceptable plan is ‘down’. Then it must be the case, by IRU, that in Figure 7b the uniquely acceptable plan is ‘down, down’. However, that implies that in Figure 7b’ the uniquely acceptable plan must be ‘down’. But it was also supposed that \( y \) is at least as good as \( z \). Then, by SR, ‘up’ in Figure 7b’ must also be acceptable. Therefore, it cannot be the case that ‘down’ is uniquely acceptable in 7b’. Then, (by SR, SEP and DC) it cannot be the case that \( n_1 \) is not part of acceptable plan in Figure 7a. Therefore, there exists an acceptable plan in 7a that calls for the continuation ‘..., up’.

Then IRU and DC require that the plan continuation ‘..., up in Figure 7c must be part of an acceptable plan as well. This in turn can only be the case if in Figure 7c’ ‘up is acceptable because of SEP. SR leads us to the conclusion that \( x \) is at least as good as \( z \). Therefore, if \( x \) is at least as good as \( y \), and \( y \) is at least as good as \( z \), the principles of dynamic consequentialism require us to conclude that \( x \) is at least as good as \( z \). Again, note that similar demonstrations can be given for alternative rankings of \( x \) and \( y \), and \( y \) and \( z \) respectively.
3. **Contraction- and expansion consistency**

The conditions of acceptable planning require that the ranking satisfies contraction and expansion consistency. Suppose, as before, that the plan ‘up, up’ is an acceptable plan in Figure 7a. IRU tells us that if the plan leading to \( x \) is acceptable in Figure 7a, \( x \) should also be the prospect of an acceptable plan in Figure 7d. Since we have already proven that the ranking is connected, we can infer (by SR) that \( x \) is at least as good as \( y \) when offered the choice between \( x, y \) and \( z \). SEP requires that the plan ‘up’ in Figure 7a' is acceptable, which (by SR) means that \( x \) is at least as good as \( y \) in a smaller set of alternatives. Therefore, the ranking is contraction consistent: if \( x \) is at least as good as \( y \) in the set \( \{x, y, z\} \), \( x \) is at least as good as \( y \) in the set \( \{x, y\} \).

It should also be the case that the reverse holds. That is, if \( x \) is at least as good as \( y \) in \( \{x, y\} \), it should not be the case that in a larger set of alternatives, \( x \) no longer is at least as good as \( y \). If \( x \) is at least as good as \( y \) in the smallest set of alternatives, it should be the case that (up) is an acceptable plan in Figure 7a'. Suppose, for reduction, it were the case that in the larger set of alternatives, \( y \) is at least as good as \( x \) but the reverse does not hold. Then if there is an acceptable plan in Figure 7a that goes through \( n_{1a} \), the only acceptable plan continuation would be (..., down). SEP then requires that (down) is the unique acceptable plan in Figure 7a', which would be contrary to what we supposed. Therefore, the ranking over the alternatives is also expansion consistent. Again, notice that similar proofs can be given for other rankings over \( x, y \) and \( z \). In conclusion, we have established that if there is a plan in Figure 7a that is acceptable, the ranking of the outcomes satisfies the requirements of ordering.
Notes

* I would like to thank Gijs van Donselaar, Maarten Franssen, Gouvert den Hartogh, Ned McClennen, Hans Rott and David Silver, as well as the various audiences to which I presented this material, for their comments and suggestions.

1 (Scheffler 1988) characterizes consequentialism more narrowly. On his definition a theory is consequentialist if it has a principle that ranks states of affairs from best to worst from an impersonal point of view, and requires agents to produce the best possible state of affairs she is in a position to produce. My definition of consequentialism is weaker for it does not require that the standard of evaluation of the consequentialist be impersonal. All that I assume is that a consequentialist is dedicated to maximizing with regards to her values.

Recently this assumption has been under attack. For example, the idea that all considerations of value can be compared has been challenged (Chang 1997). Clearly, if it is not the case that A is at least as good as B, nor that B is at least as good as A, we cannot maximize since it is impossible to determine which of these two options is the best. Another attack on the ideal of maximization comes from authors who defend *satisficing* as the correct response to values, e.g. (Slote 1989). In this paper I will abstract from both these challenges.

2 For example (Parfit 1984), (Jackson 1991) and (Broome 1991).

3 Note that I am interpreting rational choice theory as a normative theory. Many social scientists, however, think of rational choice theory as a modeling theory. They assume that agents act in ways described, represented and predicted by the theory and formulate their models and theories in accordance with this assumption. I am not addressing this use of rational choice theory in this essay, though there are some implications for this way of interpreting the theory.

4 ‘An almost unquestioned hypothesis of modern normative decision theory is that acts are valued by their consequences.’ (Hammond 1988).

5 I simplify matters here considerably. Often it is the case that the best thing *in the long run* is to let oneself be guided by rules and practices that sometimes prescribe non-maximizing actions. Many consequentialists have recognized this and argued that consequentialism should be applied to rules and practices rather than
actions in those cases. I ignore these issues of indirect versus direct consequentialism in this essay. In addition, several of the most influential consequentialists have argued that consequentialism is not about how to choose. Instead, it deals with what makes an action (or a rule, a practice, etc.) right. That is, they draw a sharp distinction between consequentialism as a theory of moral or rational truth and consequentialism as a decision procedure. They argue that only the former understanding of consequentialism is plausible. See, for example, (Parfit 1984, pp.24-29, 31-45, 98-100).

6 For a formal characterization of ordering and the related assumptions of connectedness (or comparability), transitivity and contraction- and expansion consistency, see the appendix.

7 Let $O$ be the set of available outcomes and let $P=\{x_1, y, \ldots, z\}$ be the set of prospects that have only members of $O$ as their elements. Continuity says that $\forall x, y, z \in P: (xRy \land yRz) \Rightarrow ([x, p; z, 1-p] R y \land y R [x, q; z, 1-q])$ for some values for $p$ and $q$ such that $0<q<p<1$. This axiom is basically a measurement requirement. It allows for the construction of a continuous scale, i.e., a scale that will be able to assess the value of any prospect over $O$.

8 Examples of such a numerical interval scale are the Fahrenheit and Celsius scales for temperature. See also (Broome 1991).

9 (Parfit 1984); (Jackson 1991).

10 (Broome 1991, 1999)

11 (Allais 1953).

12 To name but a few critics: independence has been criticized by (Allais 1953), (Machina 1989), (McClennen 1990) and many others. (Levi 1986) and (Loomes and Sugden 1982) have criticized ordering. (McClennen criticizes part of the ordering condition in addition to his sustained attack on independence.) A good, if somewhat dated, overview of the different proposals can be found in (Fishburn 1988).

13 (Hammond 1988).

14 (Raiffa 1968).

15 See (McClennen 1990)

16 (Rabinowicz 1995).

17 Thanks to one of the referees for reminding me of the crucial role this assumption plays in the argument.

18 These assumptions are for heuristic purposes only. Introducing the possibility of changing information over time and resource constraints, complicates matters needlessly for the present purposes.
19 I deviate from McClennen’s name for this principle here and use the term coined by (Cubbit 1996).

20 Hammond does not use dynamic consistency since he subscribes to a ‘revealed preference’ theory. On that theory, choices, rather than plans or preferences are the primitive inputs of the theory. Plans are derived from the actual sequence of choices. It follows that on Hammond’s view one can never deviate from a plan. Rather, any perceived ‘deviation’ shows that one did not hold that plan to begin with.

21 For a thorough defense of this view on plans, see (Bratman 1987).

22 In section 9 I discuss different kinds of forward-looking that could be required by consequentialism.

23 I take this to be the interpretation that (McClennen 1990) favors.

24 This is how (McClennen 1997) reformulated the condition.

25 I will not provide a general and formal proof here as that would take too much space. See (Hammond 1988) for a complete, but very technical proof. (McClennen 1990) gives a relatively simple and straightforward proof for the sufficiency of the principles of dynamic choice for independence. A more refined proof in the spirit of McClennen, can be found in (Cubbit 1996). My informal proof here as well as the one in the appendix differs from all of these because I formulate it in terms of preferences over prospects rather than choice over prospects.

26 Actually, Hammond’s assumption is slightly different. He assumes that the domain of logically possible trees is unrestricted. Suppose that a tree T has the set of outcomes \(O_T\). Hammond assumes that if there exists an acceptable plan in T, there exists an acceptable plan in all logically possible trees \(T^*\) for which \(O_{T^*} \subseteq O_T\). (Munier 1996) in his insightful comment on Hammond’s formal proof, shows that if the shape of the tree individuates the outcomes, the set of logically possible trees \(T^*\) may consist just of T. In that case, we cannot deduce independence from the existence of an acceptable plan in T. Hammond resisted this criticism for a long time. Recently, however, he has acknowledged this point. He explicitly states that his proof only applies when ‘... the consequences themselves do not depend on the structure of the tree. Of course, such independence has been the standard assumption in classical decision theory.’ [Italics added](Hammond 1996).

In a way, it is this ‘independence’ that I question in this essay.

27 An similar argument can be given if \(y\) is at least as good as \(x\) and \([x, p; z, 1-p]\) is at least as good as \([x, p; z, 1-p]\).

28 (Allais 1953)
The reflections on her choices of this hypothetical agent are representative of those of actual agents. (Tversky and Kahneman 1981) found that most subjects display this pattern of choices. The pattern was stable for varying monetary rewards and was persistent in both educated and uneducated subjects. Additional research led them to believe that human beings display all kinds of ‘certainty effects’ in their choice behavior.

These results raise some serious questions whether standard expected utility theory can be maintained as the theory of rational choice. If otherwise reasonable agents violate the prescriptions of the theory in some very systematic ways, what grounds do we have for believing that standard expected utility theory is plausible as a theory of rational choice?

(Broome 1991).

(Broome 1991).

(Broome 1991). Broome is talking here in the context of expected utility theory and the question whether one can use that theory to judge the rationality of a preference. As I have been arguing, there is nothing that restricts the principle of rational choice to the domain of preferences. The same thing applies here. Broome’s principle can very well be used to constrain value judgments in the same way as it can constrain preferences. Outcomes should be distinguished as different if and only if they differ in a way that makes it rational to judge one better than the other.

Incidentally, Brome goes on to argue that there are two logically equivalent ways in which one can use his principle. On the one hand one can use it to put restrictions on the individuation of outcomes. Alternatively, one can use it to put restrictions on the agent’s preferences. This is the most promising interpretation for the case at hand. It seems silly to deny that prospect \([$0, .01; $5M, .10]\) is a different prospect than \([$0 \text{ and regret, .01; } $5M, .10]\). Broome’s principle requires the agent to be indifferent between the two prospects if regret is not a justifier. See also (Broome 1991).

See (Broome 1999). I use these terms very loosely here to refer to the idea that there are constraints on what can count as valuable for an agent, which are in some sense independent of the agent.

Note that more subtle subjective theories like an informed preference theory of wellbeing or an ideal advisor theory of wellbeing suffer from the same problem.

See (Broome 1999).

This is what (Weber 1998) calls ‘separable outcome regret’.

38 This mirrors the conclusions of (Weber 1998) that regret or disappointment cannot rationalize the Allais preferences in the one ‘up front’ choice case of Figure 2.

39 The terminology is that of (Pettit 1997). Similar characterizations can be found in (Anderson 1993).

40 For example, (Nussbaum 1986) and (Williams 1982).

41 Note that this way of understanding Nussbaum’s and Williams’ claim is perfectly compatible with consequentialism. It is the overall expected value of a prospect that needs to be promoted. It may very well be that the best prospect contains some very bad elements. However, I doubt that this is the best way to understand Nussbaum and Williams. Their claim is that these bad elements make a contribution to the value of the prospect, and this is something that consequentialists dedicated to the expectation method will have difficulty to accommodate.

42 (Tversky and Kahneman 1981).

43 For example, (Sturgeon 1996).

44 (Pettit 1997).

45 See the appendix for the crucial role of separability in the derivation of transitivity.

46 For example, (Cubbit 1996) distinguishes between separability about past chance nodes and separability about past choice nodes. The former is necessary for the derivation of independence, whereas the latter is necessary for the derivation of ordering. On Cubbit’s view, the dynamic foundations for ordering and independence are distinct.

47 Though there are many critics of independence and ordering, the literature is less voluminous when we look at concrete alternatives for the expectation method. Within the so-called non-linear utility theory several methods to represent rankings have been proposed that do not require independence. A good example is (Quiggin 1993). (Machina 1982) has given general theoretical framework for the characterization and classification of non-linear utility theory. There are also proposals how to deal with the representation of rankings that do not satisfy ordering. Here the theory is less systematic and the few concrete proposals are more ‘data driven’. Examples are (Kahneman and Tversky 1979) and (Loomes and Sugden 1982).

48 This is equivalent to conditions $\alpha$ and $\beta$ that (Sen 1970, chapter 1*) identifies.
References


