Discussion: DeRose on the Conditionals of Deliberation

I take issue with two claims of DeRose: Conditionals of deliberation must not depend on backtracking grounds. 'Were'ed-up conditionals coincide with future-directed indicative conditionals; the only difference in their meaning is that they must not depend on backtracking grounds. I use Egan's counterexamples to causal decision theory to contest the first and an example of backtracking reasoning by David Lewis to contest the second claim. I tentatively outline a rivaling account of 'were'ed-up conditionals which combines features of the standard analysis of counterfactuals with the contextual relevance of the corresponding indicative conditionals.

Keith DeRose has recently addressed two questions which are of crucial importance to a general theory of conditionals (DeRose 2010): (i) One main function of conditionals is practical deliberation. We deliberate what the consequences are given we perform some action. But what are the conditionals suitable for expressing such deliberations? (ii) There is a fairly standard view according to which indicative and subjunctive conditionals are distinguished by the latter usually expressing counterfactuals, at least when they are of the form: 'If A had been the case, C would have been the case.' But how are we to understand future-directed 'were'ed-up conditionals ('If A were the case (at some future time t), C would be')?

DeRose answers: (i) Practical deliberation usually proceeds by indicative conditionals; yet in order to be deliberationally useful, conditionals must not depend on *backtracking grounds*. (ii) There are no genuine future-directed counterfactuals expressible by 'were'ed-up conditionals. What looks like a counterfactual, in fact roughly shares the semantics of indicative conditionals, although it slightly diverges in assertability conditions.

I want to criticize both answers in light of some counterexamples. I argue for two claims:

1

1. What disqualifies conditionals for deliberational purposes is not backtracking.

2. 'Were'ed-up conditionals are not just souped-up indicative conditionals.

1. Conditionals of deliberation may depend on backtracking grounds

DeRose points out a problem of his central hypothesis that the conditionals of deliberation are indicatives. As the well-known counterexamples to evidential decision theory show, some indicative conditionals convey links which are merely evidential but not causal. These conditionals may give rise to ineligible courses of action if used in practical deliberation. To evade this problem, DeRose provides a criterion that allows to tell apart conditionals which may be used in deliberation and conditionals which may not. The former should not depend on backtracking grounds (DeRose 2010, 28-30). An example:

> …if Sophie is deciding between going to seminary or joining the army, and knows that (even after she has heard about the connection between her career choice and the likelihood of her having the condition) her choosing to go to seminary would be very strong evidence that she has a certain genetic condition that, if she has it, will almost certainly also result in her dying before the age of 40 years, she has strong grounds to accept that, very probably

> (P) If I go to seminary, I will die before the age of 40

> Yet, as most can sense, this, plus her desire not to die young, provides her with no good reason to choose against the seminary, for she already either has the genetic condition in

question or she does not, and her choice of career paths will not affect whether she has the condition.(DeRose 2010, 22)

To DeRose, (P) is deliberationally useless because it is backtracking:

> Sophie's grounds for (P) … involve this backtracking pattern of reasoning. After provisionally making the supposition that she goes to seminary, she then reaches backward in the causal order to conditionally alter her view of what her genetic condition is (from agnostic to supposing that she (probably) has the lethal condition), to explain how that antecedent (likely) would become true, and she then conditionally reasons forward to her untimely death.(DeRose 2010, 29)

However, some recent paradigm cases presented by Andy Egan should give us pause. Egan heralds them as counterexamples to causal decision theory.

> *The Psychopath Button*
> Paul is debating whether to press the 'kill all psychopaths' button. It would, he thinks, be much better to live in a world with no psychopaths. Unfortunately, Paul is quite confident that only a psychopath would press such a button. Paul very strongly prefers living in a world *with* psychopaths to dying. Should Paul press the button? (Set aside your theoretical commitments and put yourself in Paul's situation. Would *you* press the button? Would you take yourself to be irrational for not doing so?)(Egan 2007, 97)

By Egan's lights, sound intuition has it that Paul should not press. As it seems, any reasoning that leads to this result irremediably depends on backtracking grounds. If Paul presses, he must have been a psychopath all along in order to do so; hence he will be killed. This

reasoning exactly parallels DeRose's example of Sophie, the difference being that only Paul's conditional plays a role in evaluating *the causal consequences* of the choice to be made. There are some reservations about Egan's examples. But I have not yet seen an argument that successfully counters their intuitive pull.[1]

Independently of Egan cases, there is reason to doubt the backtracking diagnosis. DeRose's main evidence is his version of Gibbard's riverboat example (Gibbard 1981, 231-232, DeRose 2010, 21-25). I shorten the case: Sly Pete is playing poker on a Mississippi riverboat. To call means to win provided you have the higher cards. Assume Sigmund does not know how the cards are distributed. What Sigmund knows is that Pete knows the cards of his opponent. Sigmund justifiedly accepts

(O) If Pete calls, he will win

Snoopy, in contrast, knows that Pete has the losing hand. So he justifiedly accepts

(Oc) If Pete calls, he will not win

As DeRose notes, when the conditionals are reported to Pete, (Oc) is useful in deliberation but (O) is not. DeRose's explanation is that (O) depends on backtracking reasoning:

If Pete plays, that will be because he has the higher card; but then of course he will win.(DeRose 2010, 29)

---

[1]   Doubts about Egan cases are expressed by Arntzenius (2008, 291), counterarguments have been forwarded by Cantwell (2010) and Weatherson (unpublished).

Judging from the Sophie case, we should expect backtracking to go as follows: Sigmund arrives at (O) by 'provisionally making the supposition' that Pete calls and then 'reaching backwards in the causal order' such as to revise his beliefs about Pete's playing dispositions. Sophie *must* use the supposition of her going to seminary as evidence for a certain causal order to arrive at (P). That's why her reasoning *does depend* on backtracking. Nothing like that has to occur in Sigmund's reasoning. Without reaching backwards in the causal order from the supposition that Pete calls, he can derive (O) from independently justified assumptions about Pete's using method M: Arrange for knowing the cards! Call precisely if you are signalled that you have the higher cards! Hence (O) does *not depend* on backtracking reasoning.[2]

What disqualifies (O) for Pete's deliberational purposes is that for (O) to be acceptable in the first place, Pete must use method M. (O) would be undermined if he were to use (O) *instead of M* to reason: 'If I call, I will win. So I should call.' In contrast, (Oc) does not exhibit this

---

[2] In spite of these shortcomings, we might reckon the 'that will be because'-template a shibboleth of useless conditionals. Indeed this template might provide some evidence against a conditional being deliberationally useful. It indicates that the conditional draws on causal facts that are 'sunk'. Yet there are counterexamples which are less demanding than Egan cases:

Should I go to the exhibition? I should go only because I appreciate the artworks for their own sake. If I go, that will be because of my snobbery and not because of my appreciating the artworks for their own sake. So I should not go.

Note that unlike Egan, one does not have to claim that a certain choice is ultimately preferable but only that these considerations play a legitimate role in deliberating action.

pattern of dependence. More generally, a conditional is useless in a deliberational process when it is acceptable only provided the deliberational process does not depend on this very conditional.

2. 'Were'ed-up conditionals are not just souped-up indicative conditionals

By DeRose's lights, 'were'ed-up conditionals coincide with indicative conditionals, the difference being that they can be used to convey that the antecedent is probably false and that they are unassertable when the corresponding indicative conditionals for their assertability depend on backtracking reasoning (DeRose 2010, 37-38).

My criticism takes three steps. (i) I outline a counterintuition which I take to show that DeRose's solution is wrong. (ii) I summarize the main evidence assembled by DeRose. (iii) I indicate an alternative way of dealing with this evidence which combines features of the standard analysis of counterfactuals with the contextual relevance of the corresponding indicative conditionals.

2.1 Problems of DeRose's reading

Why is DeRose's approach problematic? I think the immediate intuition how to deal with 'were'ed-up conditionals is to assimilate them to other subjunctive conditionals. As a consequence, the following reasoning of David Lewis' seems to apply. Lewis famously eschews backtracking counterfactuals. Yet he allows for certain exceptions triggered by suitable contextual clues which override the standard non-backtracking solution:

Jim and Jack quarreled yesterday, and Jack is still hopping mad. We conclude that if Jim asked Jack for help today, Jack would not help him. But wait: Jim is a prideful fellow. He

never would ask for help after such a quarrel; *if Jim were to ask Jack for help today, there would have to have been no quarrel yesterday*. In that case Jack would be his usual generous self. So if Jim asked Jack for help today, Jack would help him after all.(Lewis 1986, 33, my emphasis)

Nothing seems to preclude modifying Lewis'story as follows: Just replace the last sentence by

So if Jim were to ask Jack for help later today, Jack would help him after all

This is infelicitous by DeRose's lights.[3] But it seems perfectly in order.

2.2 DeRose's evidence

To appreciate DeRose's evidence, consider the problem of future Adams pairs (cf. Morton 2004). The classical Adams pair is this:

(A) If Oswald didn't shoot Kennedy, someone else did
(B) If Oswald hadn't shot Kennedy, someone else would have (DeRose 2010, 2)

In contrast, the following future-directed pair sounds inconsistent:

---

[3] Curiously DeRose accepts that 'were'ed-up conditionals might be used in this way provided the backtracking reasoning is explicit (DeRose 2010, 35 ann. 31). But I do not see how this concession can be reconciled with his overall account of their meaning and purpose: 'Were'ing-up is a device of clearly marking out conditionals as based on the right sorts of grounds to be deliberationally useful.'(DeRose 2010, 38)

Eve

'If I put Eve into situation S_1, she will sin; but if I were to put her into situation S_1, she

wouldn't sin.'(DeRose 2010, 9)

While DeRose is more cautious (DeRose 2010, 10), one may take the Eve-pair to provide

evidence against future Adams pairs. No Adams without *Eve*. DeRose's approach neatly

explains why *Eve* sounds inconsistent: 'Were'ed-up conditionals just coincide with indicative

conditionals in the relevant repects. The independent lesson to draw is that (a) *there is an*

intimate connection between an indicative and the corresponding 'were'ed-up conditional;

the indicative is not reconcilable with the contrary 'were'ed-up conditional.

Yet the riverboat example teaches an opposing lesson. While Sigmund justifiedly accepts

(O) If Pete calls, he will win,

according to DeRose, he should reject (DeRose 2010, 32)

(Ow) If Pete were to call, he would win.

Gibbard prefers a nearness analysis as it is standard for subjunctives of the 'had-would' type

(Gibbard 1981, 228-229). In contrast, DeRose maintains:

Gibbard's response is to place (O) and (Ow) on opposite sides of the great semantic divide

among conditionals. …though I agree with Gibbard that (O) seems right and (Ow) wrong

for Sigmund, the difference between the two conditionals seems slight and subtle. They

seem to mean approximately the same thing, which, together with the sense that one seems

right and the other wrong here produces a bit of a sense of puzzlement about the situation.(DeRose 2010, 33)

DeRose's account is to dissolve this puzzlement. In contrast to (O), the 'were'ed-up (Ow) is unassertable because it rests on backtracking grounds. Yet the independent lesson to draw is (b) *sometimes there is a divide between the indicative conditional and its 'were'ed-up version; one is assertable while the other is not.*

This lesson is enforced by a further argument of DeRose's. Indicative conditionals underlie a paradox (DeRose 2010, 16-17). The following reasoning seems all right:

 (1) Either the butler or the gardener did it.

 (2) Therefore, if the butler didn't do it, the gardener did.

So does the following:

(3) The butler did it.

(1) Therefore, either the butler did it or the gardener did it.[4]

But we cannot reason as follows:

 (3) The butler did it.

 (2) Therefore, if the butler didn't do it, the gardener did.

---

[4] I am not so sure whether this really seems compelling to the untutored. But let us grant the point.

While DeRose has it that future indicative conditionals underlie the paradox, he reports mixed intuitions as to whether 'were'ed-up conditionals do (DeRose 2010, 36-37 ann.). His explanation is this: We cannot simply reason

(1´) Either the butler or the gardener will do it.

(2´) Therefore, if the butler were not to do it, the gardener would.

For when we accept (1´), still we cannot be sure that the assertability conditions of (2´) are met. Backtracking reasoning might be involved. There are two misgivings: Firstly, assume we check and rule out first that any of our premises depends on backtracking; then the reasoning should seem convincing. If we are still reluctant, this calls for a different explanation. Secondly, why are the results mixed? If DeRose were right, every competent speaker should feel the same about (1´)-(2´).

2.3 How to deal with DeRose's evidence

Given my intuitions about Lewis' Jim-and-Jack example, I need a way of reconciling lesson (a) and (b) that diverges from DeRose's:

(a) the close connection between indicative and the corresponding 'were'ed-up conditionals that rules out *Eve*

(b) the difference between indicative and corresponding 'were'ed-up conditionals that accounts for

- (O) being assertable but (Ow) not

- the paradox of indicative conditionals pertaining to future-directed indicative conditionals but not undoubtedly to their 'were'ed-up version.

Ad (a) A leitmotiv of DeRose is that indicative and 'were'ed-up conditionals are too close to each other to be placed on 'opposite sides of the great semantic divide'. Yet this can be accommodated as follows: The factual/counterfactual-distinction is not as well marked with respect to the future as with respect to the past. This distinction is crucial for the sharp boundary between indicative and subjunctive conditionals as it is manifested in past-directed Adams pairs. We take the past to be fixed. Past-directed indicative conditionals are assessed by (hypothetically) taking the antecedent to be *part of the fixed past*. When they give rise to Adams pairs, the antecedent situations of past-directed subjunctive conditionals are taken to be *ruled out by the fixed past*. In contrast, we are prone to regard the future as not yet fixed. There is a tendency towards considering the antecedent of a future-directed conditional as an option that has not yet been ruled out and is not predetermined to come about either. As a consequence, the demarcation of future indicatives and counterfactuals tends to become obliterated. This is the reason why future indicative conditionals and their 'were'ed-up variants are so close to each other; and why the very same antecedent possibility that is envisaged in the indicative partner of an alleged future Adams pair like *Eve* is counted among the antecedent possibilities relevant to evaluating the contrary 'were'ed-up version. In the very same scenario of Eve being put into situation S_1, she would have both to sin and not sin for the conditionals to be reconcilable. Yet by opting for the 'were'ed-up version we express that we feel hesitant about the antecedent situation coming about in due course, i.e. in the way the indicative conditional conveys; as a consequence, we normally open the range of situations relevant to evaluating the 'were'ed-up version for the standard ways we take a counterfactual antecedent situation to come about.

There are different ways of further elaborating these findings. I want to keep my approach as simple as possible. To start with, although one should beware of going 'Into the swamp' of

indicative conditionals,[5] I need some minimal common ground between indicative and (a standard view of) subjunctive conditionals:

A conditional 'If A, C'/'If A were the case, C would be' is true/assertable iff C in all salient A-situations.

I hope that my use of this vague condition squares with DeRose's Ramseyan account: '…one is positioned to assert A→C if and only if adding A as a certainty to one's belief set would put one in a position to assert that C.'(DeRose 2010, 15) Nothing I say should preclude the situations salient in indicative conditionals from being those that vindicate one's beliefs about the actual world updated with the certainty A.

I combine this with a simplified standard nearness analysis of 'were'ed-up conditionals. Just let the salient A-situations be those that are closest or most similar to the actual situation. When we ask ourselves how the antecedent A might come about, the A-situations envisaged in the future indicative spring to mind. Drawing on the Ramsey test, I surmise that these situations are those that vindicate our belief system updated with A. We usually reckon them among the closest A-situations with regard to which the subjunctive is assessed.

I suggest the following constraint on 'were'ing-up:

Eve-constraint

Whenever a future-directed indicative conditional 'If A, C' is assertable, an A-cum-C-situation must be among the closest situations relevant to evaluating its 'were'ed-up versions.

More precisely, whenever the indicative conditional is assertable, the contrary 'were'ed-up version 'If A were the case, C would *not* be the case' is not. Yet in evaluating the 'were'ed-up

---

[5] Cf. DeRose 2010, 39-40.

conditional, we also attend to ways in which we take a standard counterfactual situation to come about; hence the A-cum-C-situations envisaged in the indicative conditional normally are *only one among several* candidates for the closest A-situations. As a consequence, the assertability of the indicative conditional normally is not sufficient for the 'were'ed-up version being assertable as well. Pace Gibbard, I think that DeRose is completely right *not* to simply place indicative conditionals and their 'were'ed-up versions on 'opposite sides of the great semantic divide'. Yet I follow Gibbard's nearness analysis. As far as the great divide exists for future-directed conditionals, it cuts through 'were'ed-up conditionals.

Ad (b) There is an eligible explanation of our rejecting (Ow): A standard Lewisian analysis is available which parallels the notorious Nixon example (Lewis 1986):

If Nixon had pressed the button, nuclear holocaust would have ensued.

Lewis proposal under determinism is this: By default, we take a small miracle to bring about Nixon pressing the button, say an additional neuron firing in his brain. Under indeterminism, some comparable chance process makes Nixon press the button. We do not resort to Nixon's reasons for pressing the button or the like.

This analysis may be transferred to (Ow): We take a small inconspicuous divergence to bring about Pete calling. We do not resort to Pete's reasons for calling or the like. So we do not care about Pete knowing the cards of his opponent and reacting rationally. Since we do not posit a connection between Pete's calling and the distribution of cards, we have no reason to assume that Pete will win in all salient situations.

However, taking into account the constraint that rules out *Eve*, we cannot simply settle for these criteria. (O) seems less clearly distinguished from (Ow) than

If Pete called, he won

is distinguished from

If Pete had called, he would have won.

as uttered by Sigmund from an ex post perspective yet given the same evidence.

This can be explained by the asymmetry between past-directed and future conditionals. Lewis' criteria are *partially overridden* by the *Eve-constraint*: The antecedent-cum-consequent-situations that are salient in (O) are among the closest situations considered in evaluating (Ow). So we tend to amend Lewis' criteria by this constraint. But this is reconcilable with our adhering to them to a certain extent. Among the closest situations considered are situations where Pete's calling comes about by a small miracle or the like. This accounts for our rejecting (Ow).

Concerning Lewis' Jack-and-Jim example, the nearness account explains why we accept

If Jim were to ask Jack for help, Jack would help him after all.

Since we tend to rule out that Jim will ask Jack, we do not feel inclined to the indicative conditionals

If Jim asks, Jack will / will not help him.

If we deem them unassertable, the *Eve constraint* is not binding. In this case, the 'were'ed-up conditional is treated according to a standard nearness analysis and converges to the counterfactual

If Jim asked for help today, Jack would help him after all.

But assume we are pressed about the indicative conditional ('yes, but *if* Jim asks?'). If we deem an indicative conditional assertable, the context has us rather accept 'If Jim asks, Jack will help him'. Then the *Eve constraint* supports the 'were'ed-up version.

A general concern: How can situations that are framed so differently, on the one hand in terms of the Ramsey test, on the other hand in terms of Lewis' small inconspicuous divergence count as equally close? Due to the specific openness we accord to the future, we waver between two quite different options for closeness: The first is to take the antecedent A as a new piece of evidence in light of which we revise our view of the actual situation. So we consider the situations that vindicate our system of beliefs about the actual world updated with A. Yet by the subjunctive mood, we signal that we do not simply take the actual situation as giving rise to A in due course. Hence we also consider the closest situation which is different from the actual course things will take. *That* closest situation is not reckoned a candidate for updating our beliefs about the actual world. It is distinguished from the actual world by a small inconspicuous divergence (miracle) that brings about A.

The remaining task is to account for the mixed results regarding the paradox of indicative conditionals. To begin with, although the reading just developed is the default reading, there may even be a reading of (Ow) in which the situations that are salient in (O) *completely override* Lewis' criteria. One may attend exclusively to the features which guide (our prediction of) Pete's rational deliberations: Pete knows the cards of his opponent, he knows the rules of the game, he aims at winning and so on. I have suggested that these are the relevant features common to situations that vindicate our belief system updated with 'Pete calls'. Under these circumstances, the closest situations in which he calls will be situations in which he wins. As a consequence, in one non-default reading 'were'ed-up conditionals *come*

close to the corresponding indicative conditionals, perhaps so close as to coincide with them. Yet we do not settle for this reading unless there are sufficient clues enforcing it.

On this basis we may account for the paradox of indicative conditionals. Those who deem the inference (1´)-(2´) invalid, treat the 'were'ed-up conditional according to a default nearness analysis modified by the *Eve constraint*. For instance, they take into account that while the gardener is innocent, the butler is about to do it but some small miracle interferes. Hence they deny that if the butler were not do to it, the gardener would. Those who tend to accept the inference follow the contextual pull of assimilating it to the corresponding indicative conditional. Where may this pull come from? (1)/(1´) focus attention on the possibilities of the butler and the gardener doing it. In order for the indicative conditional (2) to follow from (1), one must *rule out* any further possibilities as salient.[6] The contextually relevant situation in which the butler does not do it is one in which the gardener does.

The result is a neat picture of 'were'ed-up conditionals:

'Were'ed-up conditionals conform to the standard analysis of counterfactuals, the difference being that Lewis' default criteria are either

- *partially overridden by the Eve-constraint (the standard case)*

or

- *completely overridden by the situations that are salient in the corresponding indicative conditionals (given certain contextual clues).*

Arntzenius, F. (2008). 'No regrets, or: Edith Piaf Revamps Decision Theory'. *Erkenntnis* 68, 277–297.

---

[6] Otherwise one could not assert that the gardener did it upon adding that the butler did not do it as a certainty to one's belief system.

Cantwell, J. (2010). 'On an Alleged Counter-Example to Causal Decision Theory'. *Synthese* 173, 127-152.

DeRose, K. (2010). 'The Conditionals of Deliberation'. *Mind* 119, 1-42.

Egan, A. (2007). 'Some Counterexamples to Causal Decision Theory'. *The Philosophical Review* 116, 93–114.

Gibbard, A. (1981). 'Two recent theories of conditionals'. In W.L. Harper et al. (ed.), *Ifs*. Dordrecht: Reidel, 211-47.

Lewis, D. (1986). 'Counterfactual Dependence and Time's Arrow'. In *Philosophical Papers: Volume II*. Oxford: Oxford University Press, 32-52.

Morton, A. (2004). 'Indicative versus Subjunctive in Future conditionals'. *Analysis* 64, 289–93.

Weatherson, B., 'Defending Causal Decision Theory', unpublished manuscript.