Letters from Long Ago: On Causal Decision Theory and Centered Chances

Rabinowicz, Wlodek

Published in:
Logic, Ethics, and All That Jazz - Essays in Honour of Jordan Howard Sobel

2009

To Howard Sobel, with fondness and admiration

Letters From Long Ago: On Causal Decision Theory and Centered Chances[*]

Wlodek Rabinowicz

Department of Philosophy, Lund University
Wlodek.Rabinowicz@fil.lu.se

This paper goes back to a discussion that took place nearly thirty years ago, in the early 1980-ies. The participants were David Lewis, Howard Sobel and I. The background of the exchange was as follows. In Lewis (1981),[1] Lewis presented his own version of causal decision theory and went on to argue that the differences between this version and the other versions – the ones that had been put forward by Gibbard and Harper, Skyrms and Sobel – were relatively negligible. As he put it, causal decision theorists 'differ mainly in matters of emphasis and formulation' (ibid., p. 5 [306]). As a young and enthusiastic adherent of causal decision theory I was at that time much influenced by the seminal work of Gibbard and Harper (Gibbard & Harper 1978), and by Sobel's complicated and subtle *Probability , Chance and Choice* (Sobel 1978 [1980]). Lewis was one of my philosophical heroes, but I found his attitude to the different proposals in the area too ecumenical, which I tried to show in 'Two Causal Decision Theories: Lewis vs Sobel' (Rabinowicz 1982a). Upon receiving a copy of that paper, Lewis sent me a very nice letter (Lewis 1982), in which he accepted some of my critical points, rejected others and further clarified his standpoint. In particular, he clarified his views on the connection between subjunctive conditionals and statements about chances. Some of the observations he made in that letter were afterwards presented in print, in Lewis (1983).

I responded with a long letter of my own (Rabinowicz 1982b), to which, I am sorry to say, I never received an answer. Lewis' patience with the young critic must have run out at that point! The years passed, I forgot the whole exchange and only got reminded about it when Stephanie Lewis kindly made available to me a scanned file with this correspondence. The file contains both Lewis' letter (three typewritten pages) and my own response (thirteen handwritten ones). There is also a short message from Sobel (Sobel 1982), addressed to both of us, to which he apparently attached his comments. (He had copies of both our letters.) Unfortunately, the comments themselves are missing from the file.

Having now re-read that correspondence, it seems to me that one of the ideas in my letter to Lewis might be worth reviving. I have in mind the suggestion that causal decision theory should be formulated in terms of *centered chances*. The expected utility of an option *A* may be seen as a weighted sum of the utilities of *A* at different possible worlds, with weights being the credences that the agent assigns to these worlds. The utility of *A* at a given world is interpreted as a weighted sum of the values of *A*'s different possible outcomes, with weights

[*] I am grateful to Huw Price for the impulse to write this paper, to Stephanie Lewis for the access to my correspondence with David Lewis and to Staffan Angere, John Broome and Frank Zenker for valuable suggestions. My work on this paper started in the Fall of 2008, while I was staying at the Swedish Collegium of Advanced Study in Uppsala. I am indebted to the Collegium for its hospitality and to the Swedish Research Council for a generous research grant.

[1] This paper was re-printed in Lewis' *Philosophical Papers*. In what follows, page references are to the original text, with the references to the re-print within square parentheses.

being the chances (in that world) of these outcomes if *A* were performed. On the centered-chance view, the chances to be used as weights in the definition of utility are centered. Unlike ordinary chances, centered chances depend not only on what happens prior to the agent's choice but also on the events that take place after the choice. It is such centered chances that on this view are relevant to the actual utility of an action. Thus, to give an example, suppose that the action under consideration results in a bad outcome due to some event whose ordinary (i.e. non-centered) chance of occurring was very low at the time of choice. Then the utility of that action in the actual world could be high on the non-centered view, but on the centered view that utility is negative (as distinct from its expected utility), since the centered chance of the event in question given the action was one, given that it did actually take place. The resulting theory is, in my opinion, philosophically more satisfactory than the extant proposals, even though it doesn't differ much in its practical recommendations, with the exception of some rather peculiar cases.

The idea of using centered chances in the definition of utility had originally been floated, but not advocated, by Sobel (1978 [1980]), section 6.42. He became much more positive to that idea later on,[2] at the time when I wrote my letter to Lewis. But – if my memory doesn't fail me – he subsequently gave it up. I did not, but I never tried to develop this proposal in print. In case other people might find the idea attractive, I decided to reproduce the relevant parts of my correspondence with Lewis, without making any changes, apart from correcting the typos and adjusting formal notation. Clarifications and comments are provided in footnotes.

To prepare the ground, however, I first need to describe and compare Lewis' and Sobel's versions of causal decision theory. The excerpts from the correspondence are presented afterwards. In the concluding section, I provide my present views on the issue under debate. In particular, I suggest how the concept of centered chances could be defined and examine the implications of employing that concept in spelling out the consequentialist position in ethics. As I also point out, centered chances have other uses as well. Thus, the connection between subjunctive conditionals and chances can be preserved, if chances are taken to be centered. This move also allows for a strengthening of the connection between credence and the subjective expectation of chance.

1. Lewis' proposal[3]

Let a credence function *C* represent the agent's *beliefs*. We take *C* to be a probability distribution on the set *W* of possible worlds, which - following Lewis - we assume to be finite, to keep things mathematically simple. *C* is extendable to propositions, which we identify with subsets of *W*: For every proposition *X*, $C(X) = \Sigma_{w \in X} C(w)$. If $C(X) > 0$, then the conditional credence for *Y* given *X*, $C(Y/X)$, is defined as the ratio $C(XY)/C(X)$ (with *XY* standing for the conjunction of propositions *X* and *Y*). If $C(X) = 0$, $C(Y/X)$ is not defined. In case $Y = \{w\}$ for some world *w*, we shall use the notation $C(w/X)$ instead of the more cumbersome $C(\{w\}/X)$.

The agent's *desires* are represented by a value function *V* on possible worlds. $V(w)$ measures 'how satisfactory it seems to the agent for *w* to be the actual world' (Lewis 1981, p. 6 [306]). The value function is assumed to be invariant under affine transformations, i.e. to represent the agent's desires uniquely up to the choice of unit and zero. It is extendable to the expected value function on propositions. $V(X)$ – the *expected value* of a proposition *X* – is

[2] This is based on my memories of his unpublished 'Notes on some recent changes in the theory of chance and the theory of probable chance' (February 21, 1982) and of his letters to me, one of which I quote in my reply to Lewis. Unfortunately, I haven't saved this correspondence.

[3] Thi section and the two sections that follow draw on Rabinowicz (1982).

defined as a weighted sum of the values of different possible worlds, with weights being the conditional credences of these worlds given X:

$$V(X) = \Sigma_w C(w/X)V(w). \qquad (1)$$

Note that $V(X)$ is defined only if $C(X) > 0$.

The agent's options, $A_1, A_2, \ldots, A_n$, can be thought of as propositions that together form a partition of W: The agent has to decide which of them to make true. According to the classical evidentiary decision theory, the agent should choose an option A that has the maximal expected value. This advice is rejected by causal decision theory. The advice leads astray in 'Newcomb-like' problems. A might have a high expected value if performing it bears good news about the world. But this doesn't show that this option would itself be instrumental in bringing about the good results.

Lewis' version of causal decision theory is based on the notion of a *dependency hypothesis*, which is defined as 'a maximally specific proposition about how the things [the agent] cares about do and do not depend causally on his present actions' (Lewis 1981, p. 11 [313]). Whether a particular dependency hypothesis holds is independent of the agent's choice of action. Competing dependency hypotheses, $K_1, K_2, \ldots, K_m$, form a partition of W. 'Exactly one of them holds at any world, and it specifies the relevant relations of causal dependence that prevail there.' (*ibid.*) Typically, the agent doesn't know which of the dependency hypotheses K actually obtains: His credences are spread over several K's. The *expected utility* of an option A is defined by Lewis as a weighted sum of the expected values of propositions AK, for the competing dependency hypotheses K, with weights being the credences of K's:

$$U(A) = \Sigma_K C(K)V(AK) \qquad (2)$$

Lewis' theory recommends the agent to choose the option with the highest expected utility. If performing A brings good news about the world but is not instrumental in bringing about the good results, then this 'news-value' of A comes from that option's evidentiary bearings with respect to different K's. These bearings are screened-off, however, if we consider the expected values of the conjunctions AK, as it is done in (2), instead of the expected value of A alone. Therefore, the mere evidentiary bearings of A will have no influence on $U(A)$.

It follows from (1) that the values $V(AK)$ in (2) are defined only if $C(AK) > 0$. For $U(A)$ to be defined, $C(AK)$ must be positive at least for every K such that $C(K)$ is positive. Lewis has to assume, therefore, that the agent is never absolutely certain of the falsity of any such conjunction AK. 'Absolute certainty is tantamount to firm resolve never to change your mind no matter what, and that is objectionable.' (Lewis 1981, p. 14 [316]).

To prepare for the comparison with Sobel's version of causal decision theory, note that, by (1),

$$V(AK) = \Sigma_w C(w/AK)V(w) \qquad (3)$$

We know that $C(K) = \Sigma_{v \in K} C(v)$. If we let K_v stand for the dependency hypothesis that obtains in a world v, and re-write $V(AK_v)$ in accordance with (3), then (2) implies:

$$\begin{aligned} U(A) &= \Sigma_K C(K)V(AK) \\ &= \Sigma_v C(v)V(AK_v) \\ &= \Sigma_v C(v)\Sigma_w C(w/AK_v)V(w) \qquad (4) \end{aligned}$$

Following Lewis, let us at this point introduce the notion of an imaging function.[4] A function I that takes as its arguments world-proposition pairs is an *imaging function* iff it assigns to

[4] This notion was introduced in Gärdenfors (1982).

every such pair (v, X) for which it is defined a probability distribution on W that obeys the condition: $I_{v, X}(w) > 0$ only if $w \in X$.

An example of an imaging function is what might be called the *dependency function*, which to every pair (v, X) such that $C(XK_v) > 0$ assigns the probability distribution $C(\cdot/XK_v)$ on possible worlds. This means that Lewis' formula (4) for expected utility is a special case of a more general schema:

$$U(A) = \Sigma_v C(v) \Sigma_w I_{v, A}(w) V(w) \qquad (5)$$

We obtain (4) from (5) if we stipulate that the imaging function I in (5) is the dependency function. Other specifications of I will give us alternative definitions of expected utility.

2. Sobel's proposal

The *tendency function*, T, is an imaging function with the following intuitive interpretation: For all world-proposition pairs (v, X) for which T is defined and for all worlds w, $T_{v, X}(w)$ specifies the *conditional tendency* – or, in other words, the *conditional chance* – that obtains in v of the world w being the case if X were (or had been) the case. T is an imaging function: $T_{v, X}$-values for different worlds are all non-negative and sum up to 1, and they are positive only for X-worlds. T is extendable to propositions in the standard way: $T_{v, X}(Y) = \Sigma_{w \in Y} T_{v, X}(w)$. $T_{v, X}(Y)$ specifies the conditional chance that obtains in v of Y being the case if X were (or had been) the case. Note that $T_{v, X}$ might well be defined even if X is not the case in v. Even then, it will often still be meaningful to ask about the conditional chances of various outcomes Y if X had been the case.

We get Sobel's formula for expected utility from (5) above, if we take the imaging function in that schema to be the tendency function:

$$U(A) = \Sigma_v C(v) \Sigma_w T_{v, A}(w) V(w) \qquad (6)$$

The inner sum in (6), $\Sigma_w T_{v, A}(w) V(w)$, can be seen as the *utility* of an option A in a world v : It is a weighted sum of the values of different worlds, with weights being the conditional chances that obtain in v of these worlds being the case if A were performed. On this reading, the *expected* utility of an option A, $U(A)$, is a weighted sum of A's utilities in different possible worlds, with weights being the agent's credences for these worlds.

Formula (6) requires that the tendency function $T_{v, A}$ is defined for all worlds v such that $C(v) > 0$. This must hold for every option A, if comparisons of expected utility between different options are to be possible.

It should be pointed out that Sobel provides truth-conditions for subjunctive conditionals in terms of conditional chances. Thus, suppose that $T_{v, x}$ is defined. Then:

The would-conditional 'If it were that X, then it would be that Y' is true in v iff $T_{v, X}(Y) = 1$.

The could-conditional 'If it were that X, then it could be that Y' is true in v iff $T_{v, X}(Y) > 0$.

There is also room for a quantified conditional construction,

'If it were that X, then it could be, to degree k, that Y', which is true in v iff $T_{v, X}(Y) = k$.

To relate this account of subjunctive conditionals in terms of conditional chances to the Lewis-style similarity semantics for conditionals, we can use T to define a triadic similarity relation between worlds:

w is at least as similar to v as u is iff for every proposition X such that $w \in X$,
if $T_{v, X}(u) > 0$, then $T_{v, X}(w) > 0$.

Sobel imposes an appropriate restriction on T which guarantees that,

(i) For every *v*, the worlds in *W* are weakly ordered with respect to their similarity to *v*;

and

(ii) For all worlds *v*, *w* and all propositions *X*, if *T* is defined for (*v*, *X*), then $T_{v, X}(w) > 0$ iff *w* is an *X*-world that is at least as similar to *v* as any other *X*-world. [5]

Further restrictions on *T* would give us additional conditions on subjunctive conditionals. Thus, suppose, for simplicity, that for all *v* and *X* such that $v \in X$, *T* is defined for (*v*, *X*). Now, consider the following possible restrictions on *T*:

T is *weakly centered* iff, for all *v* and *X*, if $v \in X$, $T_{v, X}(v) > 0$.

I.e., equivalently, *T* is weakly centered iff every world is at least as similar to itself as every other world is.

T is *centered* iff for all *v* and *X*, if $v \in X$, $T_{v, X}(v) = 1$.

This implies that *T* is centered iff the similarity relation on worlds is centered, i.e., iff every world is more similar to itself than every other world is.

T is *fully determinate* iff for all (*v*, *X*) for which *T* is defined, $T_{v, X}(w)$ equals either 1 or 0, for all worlds *w*.

This implies that *T* is fully determinate iff for every (*v*, *X*) for which *T* is defined, there is an *X*-world that is more similar to *v* than every other *X*-world is. If, for all *v*, $T_{v, X}$ is defined for all subsets *X* of *W*, *T* is fully determinate iff for every *v*, the worlds in *W* are linearly ordered with respect to their similarity to *v*. Centering is implied by full determinacy plus weak centering, but the opposite implication does not hold: A tendency function may be centered without being fully determinate.

It is easy to see that, on Sobel's account of subjunctive conditionals, the following holds:

If *T* is weakly centered, then for all *X* and *Y*, if *XY* is true, then it is true that if it were that *X*, it could be that *Y*.

If *T* is centered, then for all *X* and *Y*, if *XY* is true, then it is true that if it were that *X*, it would be that *Y*.

If *T* is fully determinate, then for all *v* and *X* for which *T* is defined, the conditional excluded middle holds in *v*:
Either if it were that *X*, then it would be that *Y*, *or* if it were that *X*, then it would be that $\neg Y$.

Sobel thinks it is very plausible to postulate that *T* is weakly centered and highly *im*plausible to assume that *T* is fully determinate. Concerning centering, he is less categorical in rejecting this condition, but he still abstains from accepting it. To be sure, centering does have its attractions, as opposed to full determinacy: For factual suppositions, there is always a fact of the matter as to what will happen if the supposition is true, while for counterfactual suppositions, one can often only say what *could* have happened if such a supposition had been true, but not what *would* have happened under these circumstances. Nevertheless, opting for a systematic asymmetry between factual and counterfactual subjunctive conditionals is problematic in Sobel's opinion:

> My present view is that while subjunctive views that are fully centered but not fully determinate may be reasonable, they are neither mandatory [n]or exclusively reasonable. A more even-handed position

[5] The restriction on T which guarantees (i) and (ii) is as follows:

For all *XY*-worlds *v* and *w*, if $T_{v, X}(Y) > 0$ and $T_{v, X}(w) = 0$, then $T_{v, Y}(w) = 0$ provided that *T* is defined for (*v*, *Y*).

> regarding true and false antecedents is, I think, also reasonable. I am *inclined* to think that *only* even-handed positions are reasonable, but rest content with the reflection that these extremes as well as intermediary variations are accommodated in my framework. (Sobel 1978 [1980], section 6.42)

3. Comparison between Sobel and Lewis

Returning now to the comparison between Lewis' and Sobel's definitions of expected utility, let S_v be the set of all worlds that have the same tendencies as world v with respect to all the options. I.e., S_v is the set of all w such that for all options A, $T_{v,A} = T_{w,A}$ (and T is defined for (v, A) iff it is defined for (w, A)). We shall refer to S_v as the *tendency proposition* that obtains in v. Now, Lewis suggests that Sobel's tendency propositions can be identified with dependency hypotheses:

$$\text{For all worlds } v,\ S_v = K_v. \qquad (7)$$

Given (7), Sobel's formula (6) for expected utility would reduce to Lewis' formula (4) provided we make one additional assumption, namely, that

$$T_{v,A} = C(\cdot/AS_v), \text{ for all worlds } v \text{ and options } A.^{6} \qquad (8)$$

I.e., we obtain the desired reduction, if we interpret conditional chances of various outcomes if an option A were performed in terms of the agent's credences for these outcomes conditioned on the conjunction of A with the tendency proposition that holds in a given world. (8) allows us to transform Sobel's formula (6) for expected utility, $U(A) = \Sigma_v C(v)\Sigma_w T_{v,A}(w)V(w)$, into

$$U(A) = \Sigma_v C(v)\Sigma_w C(w/AS_v)V(w) \qquad (9)$$

We then get Lewis'formula (4) if we in (9) replace S_v by K_v, in accordance with (*7*).

Sobel is not prepared to accept formula (8). But – Lewis points out – Sobel's reservations about (8) are marginal: They 'entirely concern the extraordinary case of an agent who thinks he may somehow have foreknowledge of the outcomes of chance processes.' (Cf. Lewis 1981, p. 18 [321].) About such abnormal cases, Lewis himself no firm views, he writes, but he considers them to be 'much more problematic for decision theory than the Newcomb problems'. (*ibid.*)

What kind of abnormal cases Lewis has in mind? How can it be that $T_{v,A} \neq C(\cdot/AS_v)$? A case of this kind might look as follows. Suppose the agent believes that an option A at his disposal would initiate a chance process which equally well might or might not eventuate in an outcome Y. More precisely, suppose that the agent concentrates all his credence on worlds v such that $T_{v,A}(Y) = 1/2$. At the same time, however, the agent has a hunch that Y *will* in fact occur if A is going to be performed. I.e., his conditional credence $C(Y/A)$ is higher than 1/2. Since tendency propositions form a partition of W, it is easy to see that $C(Y/A) = \Sigma_S C(S)C(Y/AS)$, where S varies over tendency propositions. Therefore, since $C(Y/A) > 0$, there must exist some v to which the agent assigns positive credence and for which $C(Y/AS_v) > ½$. But then, for these worlds v, $C(Y/AS_v) > T_{v,A}(Y)$, contrary to what is stipulated in (8).

If Y is a desirable outcome, then $U(A)$ will be higher on Lewis' formula for expected utility than on Sobel's. Consequently, it might turn out that A is suboptimal on Sobel's view, but at the same time optimal on Lewis'.[7]

---

[6] There is a close connection between this assumption and Lewis' famous Principal Principle. For the latter, cf. Lewis (1980), and for the connection see Lewis (1981), pp. 27f [334f].

[7] Note, howeer, that the way in which this example is described presupposes that the tendency function is *not* centered. Only under this presupposition the agent can be certain that the tendency for Y given A is 1/2 and yet consider Y to be more likely than not on the assumption that A will be performed.

But how can such a situation arise? How can one take oneself to have a foreboding of the outcome of a chance process? Sobel writes:

> One possibility is that [the agent] believes that he has 'direct present access' to *Y*, that he is with respect to *Y* clairvoyant or a seer. Another is that he believes that *Y* has been 'revealed', that he has it on some unimpeachable authority that *Y*. […] But can beliefs in clairvoyance and revelation be reasonable? […] We leave unanswered these questions […] (Sobel 1978 [1980], section 6.71)

In Rabinowicz (1982a), I argue that Lewis cannot defend his reduction of conditional chances to conditional credences by simply dismissing forebodings and premonitions on the grounds of their irrationality: According to Lewis, decision theory should be able to give advice to the agents that is based on their own beliefs and desires, whether these beliefs and desires happen to be rational or not. He makes this claim in the course of his argument that decision theory cannot require the agents to possess full self-knowledge. (Cf. Lewis 1981, p. 10 [312]) But the point generalizes to other failures of rationality, such as possession of unfounded beliefs. As we shall see, however, from his letter cited below, Lewis himself has no intention to dismiss premonitions on these grounds: He thinks it is logically possible for premonitions to be reliable information sources.

In my paper from 1982, I also present a number of other difficulties for Lewis' reduction of Sobel's theory to his own proposal. Here are some of them:

(i) Lewis' reduction formula (7), according to which $S_v = K_v$ for all v, cannot be correct: Tendency propositions cannot be identical with dependency hypotheses, simply because they are much more specific. A dependency hypothesis spells out how things the agent *cares about* do and do not depend on his options. But, 'when it comes to the tendency propositions, all things count, even those that the agent does not care about.' (Rabinowicz 1982, p. 311). This means that while a tendency proposition is always included in some dependency proposition, the opposite need not hold: A dependency hypothesis can be partitioned into several tendency propositions.[8]

(ii) Lewis' reduction formula (8), according to which $T_{v, A} = C(\cdot/AS_v)$, leads to an obvious problem. Sobel might insist that $T_{v, A}$ can be well-defined even in some cases in which $C(\cdot/AS_v)$ is undefined, because $C(AS_v) = 0$. It doesn' help to follow Lewis' lead and require the agent to be open-minded in his credence assignments to conjunctions AS_v. $T_{v, A}$ could be well-defined even if AS_v is an empty set, in which case, of course, $C(AS_v)$ *must* equal 0. If $AS_v = \varnothing$, the agent cannot perform A in v, but it could still be meaningful to inquire what would the chances of various outcomes be if he *did* perform A. AS_v might perhaps be empty even for some worlds v to which the agent assigns positive credence. Under these conditions, formula (9) will not be applicable, in contrast to Sobel's formula (6). It seems clear, therefore, that Lewis' reduction proposal is not satisfactory.

(iii) But what if we deny that AS_v can be empty if $C(v) > 0$? If A is an option that is available to the agent, then it is arguable that he must be able to perform it in every world to which he assigns positive credence. This would allow us to answer objection (ii) above, provided we also require the agent to be open-minded. But then we would get another problem instead: It would not be possible for the tendency function T to be centered but not fully determinate with respect to the agent's options. More precisely, the following holds (cf. Rabinowicz 1982a, p. 313):

[8] As we shall see in the next section, this objection can be dealt with. The two objections that follow are, however, more serious.

Theorem 1: If T is centered and $AS_v \neq \varnothing$, then for every world w, $T_{v,A}(w)$ equals 1 or 0.[9]

Consequently, if T is centered and AS_v is non-empty for *all* the options A, then T is fully determinate in v with respect to *all* the options. This means that the assumption of non-emptiness cannot be made by someone who is attracted to centering but finds full determinacy unacceptable.

This should suffice as a presentation of the background of my correspondence with Lewis. In what follows, I present excerpts from Lewis' letter (section 4) and from my own reply (section 5). Passages that are too technical or concern issues on which I haven't focused in this paper are omitted. I have corrected some typos, adjusted page references and changed formal notation to make it more consistent with the one I have been using above. Clarifications and comments are provided in footnotes.

4. Dear Rabinowicz

Thank you very much for your paper 'Two Causal Decision Theories'. I've read it with great interest, and it has increased my understanding of the relation between my treatment and Sobel's. Let me comment on the several of the points you raise.

Concerning centering. Certainly I am committed to some form of centering: if A and B hold at world w, then $A \mathrel{\square\!\!\rightarrow} B$ holds at w, and that is because no other A-world is as close to w as w itself is.[10] But if A and B hold in w, I do not want to conclude that, at w, there is a tendency of 100% to w and hence to B, and zero tendency to any other world, given A. At least, this is so if I follow Sobel's lead and understand that 'tendency' is to be related to our ordinary talk about chance. For instance, suppose a coin is shortly to be tossed. Suppose that in fact it will fall heads. If it were tossed, it might with 50% chance fall heads; if it were tossed, it might with 50% chance fall tails. (For suppose it is a fair coin in an indeterministic world.) But if it were tossed, it *would* fall heads – for it will be tossed and will fall heads – and it is not so that if it were tossed, it might fall heads. In Sobel's terminology, I think I should say that at this world, there is 50% tendency for it to fall heads if tossed, 50% tendency for it to fall tails if tossed.

So, if 'tendency' connects to ordinary talk of chance, and if I am to hold on to centering for the similarity relation that governs would- and might-counterfactuals, then I cannot also hold on to the connection between that similarity relation and tendency that you state as (ii) on page [305].[11] So let that connection be cut. Then I have a disagreement with Sobel (1978 [1980]) about would-counterfactuals;[12] but not a disagreement which affects our decision theories. I try to assimilate Sobel's tendency function to my imaging function given by

---

[9] *Proof of Theorem 1:* Suppose that T is centered and $AS_v \neq \varnothing$. We need to show that for all w, $T_{v,A}(w)$ equals 1 or 0. Since $AS_v \neq \varnothing$, there is some u such that u belongs to AS_v. Since T is centered, $T_{u,A}(u) = 1$. Since u belongs to S_v, $T_{v,A} = T_{u,A}$. Consequently, $T_{v,A}(u) = 1$. But then, since $T_{v,A}$ is a probability distribution, it follows that for all w, $T_{v,A}(w)$ equals 1 or 0. □

[10] $A \mathrel{\square\!\!\rightarrow} B$ stands the would-conditional: 'It it were that A, then it would be that B'.

[11] Page references in square parentheses are to Rabinowicz (1982a). Lewis himself is referring to page numbers in the draft I sent him. Condition (ii) he mentions states that $T_{v,X}(w) > 0$ iff w is an X-world that is at least as similar to v as any other X-world. (See section 2 above.) Since Lewis provides a semantics for counterfactuals in terms of a similarity relation between worlds, which he takes to be centered (in the sense that every world is more similar to itself than every other world), condition (ii) would imply that T is centered as well (in the sense that for every X-world v, $T_{v,X}(v) = 1$), which he wants to deny.

[12] As we have seen above (section 2), Sobel takes it that a would-counterfactual, 'If it were that X, then it would be that Y', is true at v iff the $T_{v,X}(Y) = 1$. Since Lewis rejects centering for tendencies but accepts it for would-counterfactuals, he must reject Sobel's account of the latter in terms of the former.

$$I_{v,A}(w) = C(w/AK_v),$$

and that function will, on my view, not be (more than weakly) centered (in the sense of your page [306]) in all cases. That is, it may happen that $C(v/AK_v) \neq 1$. In the sense of centering relevant to use of imaging or tendency functions in calculating utility, I think I join Sobel in centering only weakly.

I think this may answer the difficulties posed by Theorems 1 and 2 on page [313],[13] since I think these use the centering of the imaging function that I would not accept.[14]

Concerning dependency hypotheses. Let a *practical dependency hypothesis* be 'a maximally specific proposition about how the things [the agent] cares about do and do not depend causally on his present actions' – that is, let it be what I called simply a 'dependency hypothesis'.[15] Let a *full dependency hypothesis* be a maximally specific proposition about how all things do and do not depend causally on the agent's present actions. Then I agree completely with your observation on page [311] that a tendency proposition (an equivalence class under the relation of having the same tendencies) is, if anything, not a practical but a full dependency hypothesis. However, this will not lead to further differences between the treatments if, as I think, I could just as well have formulated mine throughout in terms of full rather than practical dependency hypotheses. The difference in formulations won't matter if, whenever K is a practical dependency hypothesis and H is a full dependency hypothesis compatible with – and hence implying – K, then $V(AH) = V(AK)$. And that will be so if, for every value-level proposition $V = k$,[16] $C([V = k]/AH) = C([V = k]/AK)$. And how can that not be so if, indeed, K covers all respects of dependence that the agent cares about?[17]

Concerning agents who think they may have foreknowledge about the outcomes of chance processes. [… Let me call these cases *abnormal*.] Sobel doesn't want to reject the possibility that a fairly reasonable agent might find himself in an abnormal case. Neither do I. So far, no disagreement. But we handle the problem differently. I impose the principle (restricted to the agent's options) [to the effect that, for all options A and worlds v, $T_{v,A} = C(\cdot/AS_v)$] and accordingly agree to set aside 'some very special cases – cases about which I, at least, have no firm views'. [Cf. Lewis (1981), p. 18 [321].] Sobel doesn't impose the principle and doesn't set aside the abnormal cases. But – unless Sobel has some other reason than he gave […] – we seem to be in agreement both (1) in our treatment of the normal cases, and (2) in our hesitancy to commit ourselves to much concerning the abnormal cases.

Now, I don't dismiss the abnormal cases for the reason you consider on page [316]: that an agent would have to be irrational to be in such a case. As you say, I want the theory – so far as possible – to apply to imperfectly rational agents. Besides, I believe in the logical possibility

---

[13] For Theorem 1, see the preceding section. As for my Theorem 2, I think Lewis here refers not to the theorem itself but to a lemma I prove in the course of proving that theorem:

Lemma: If T is centered, then AS_v is at most a unit set, for all v and all options A.

Proof of Lemma: Suppose that w, $u \in AS_v$. By centering, $T_{w,A}(w) = T_{u,A}(u) = 1$. Since w, u belong to the same tendency proposition S_v, $T_{w,A}(u) = T_{u,A}(u)$. Therefore, since $T_{u,A}(u) = 1$, $T_{w,A}(u) = 1$. But then both $T_{w,A}(u)$ and $T_{w,A}(w)$ equal 1. Since $T_{w,A}$ is a probability distribution on the set of possible worlds, it follows that $w = u$. □

[14] Is Lewis' reply to my objection satisfactory? The objection was that his reduction of $T_{v,A}$ to $C(\cdot/AS_v)$ implies that we cannot have centering for T without full determinacy. Now, Lewis responds to this by pointing out that T is not centered on his view (and that he needn't treat T as centered, since he rejects Sobel's connection between would-conditionals and tendencies). But I think the problem still stands. As we have seen (see section 2 above), Sobel adheres to the view that it may be reasonable (though not mandatory or exclusively reasonable) to assume centering for tendencies and reject full determinacy. But Lewis' reduction proposal would imply that this potentially reasonable view is *incoherent*. This looks like a good reason for rejecting the proposed reduction.

[15] See section 1 above and Lewis 1981, p. 11 [313].

[16] By '$V = k$' Lewis means a proposition true in exactly those worlds w for which it holds that $V(w) = k$.

[17] This is a bit cryptic, but a simpler account is provided in my reply to Lewis (see the next section).

of time travel, precognition, etc., and I see no reason why suitable evidence might not convince a perfectly rational agent that these possibilities are realized, and in such a way as to bring him news from the future. My worry is a different one. It seems to me completely unclear what conduct would be rational for an agent in such a case. Maybe the very distinction between rational and irrational conduct presupposes something that fails in the abnormal case. You know that spending all you have on armour would greatly increase your chances of surviving the coming battle, but leave you a pauper if you do survive; but you also know, by news from the future that you have excellent reasons to trust, that you will survive. (The news doesn't say whether you will have bought the armour.) Now: is it rational to buy the armour? I have no idea – there are excellent reasons both ways. And I think that even those who have the correct two-box'ist intuitions about Newcomb's problem may still find this new problem puzzling. That is, I *don't* think that the appeal of not buying the armour is just a misguided revival of V-maximizing intuitions [i.e. intuitions in line with evidentiary decision theory – the one-box'ist intuitions about Newcomb's problem] that we've elsewhere overcome.
[…][18,19]

5. Dear Professor Lewis

Thank you very much for your most interesting letter. I agree with many of your remarks. In particular, I now realize that:

(1) I should have made it clear to the reader that you cannot accept Sobel's definition of (the truth condition for) subjunctive conditionals in terms of tendencies. This is so, since you ascribe centering to the former but not to the latter.

(2) I should have noticed that the difference between 'practical' and 'full' dependency hypotheses is irrelevant as far as the calculation of expected utility is concerned. Thus, even though your assumption

$$\text{for any world } v,\ K_v = S_v,$$

Is incorrect, you could have replaced it by the obviously correct claim that,

$$\text{For any } v,\ S_v \subseteq K_v \text{ and } V(AK_v) = V(AS_v).^{20}$$

The effect, in terms of expected utility, would be the same. You would still be able to derive from your definition of expected utility,

$$\text{(i)}\quad U(A) =_{\text{df}} \Sigma_K C(K)V(AK),$$

the equality:

---

[18] Here, I omit a lengthy passage dealing with Theorem 4 in Rabinowicz (1982a). The theorem states that there exists a model – a set of possible worlds *W*, a tendency function *T* on that set and a set of options *A* - such that *no* 'normal' credence function *C* on *W* can satisfy Lewis reduction formula (8), according to which $T_{v,A} = C(\cdot/AS_v)$, for all options *A* and worlds *v*.
(*C* is 'normal' iff it assigns positive values to all non-empty subsets of *W*, i.e., iff it is a credence function of an agent who keeps an open mind with respect to all propositions that are not impossible.)
Lewis in his letter questions whether the model I use to prove the theorem makes sense. As he points out, this model allows chances themselves to be chancy, so to speak: It contains worlds *v*, *w* and *u* such that $T_{v,A}(w) > 0$, $T_{v,A}(u) > 0$, but for some proposition *X*, $T_{w,A}(X) \neq T_{u,A}(X)$. Lewis finds this absurd. To avoid such a possibility, he suggests we should impose the following restriction on all admissible models: If $T_{v,A}(w) > 0$, then $T_{v,A} = T_{w,A}$. Note, however, that if we allow *T* to be centered, what Lewis finds absurd will be a common occurrence. We shall have lots of cases in which for some $\neg A$-world *v*, there will be an *AX*-world *w* and an $A\neg X$-world *u*, such that $T_{v,A}(w) > 0$ and $T_{v,A}(u) > 0$. Given that *T* is centered, we'll then have $T_{w,A}(X) = 1$, while $T_{u,A}(X) = 0$.
[19] © Stephanie R. Lewis, with permission.
[20] More perecisely, what we need is just the following: For all *v*, $V(AK_v) = V(AS_v)$.

(ii) $U(A) = \Sigma_v C(v)\Sigma_w C(w/AS_v)V(w)$.

($\Sigma_K C(K)V(AK) = \Sigma_v C(v)V(AK_v) = \Sigma_v C(v)V(AS_v) = \Sigma_v C(v)\Sigma_w C(w/AS_v)V(w)$.)

And (ii), together with your assumption that $T_{v,A} = C(\cdot/AS_v)$ for all *v* and *A*, implies the Sobelian formula:

(iii) $U(A) = \Sigma_v C(v)\Sigma_w T_{v,A}(w)V(w)$.

(3) I should have noticed that you don't intend to *apply* your decision theory to 'abnormal' cases (cases in which the agent takes himself to have foreknowledge about the outcome of a chance process). Thus, it is wrong to suggest (as I have done on page [316]) that your theory and Sobel's might generate conflicting prescriptions in abnormal cases. Instead, what I should have said is that your theory, contrary to Sobel's, is not meant to give any guidance in cases like that.

[...][21]

Now, I turn to the issue that seems to me to be crucial when one compares your theory with Sobel's. I now think that the main difference between these two theories consists in that they express different attitudes to centering. You assume that tendencies are not centered. Sobel, in his work from 1978 [1980], does not assume centering either. But, and here comes the difference, neither does he want to assume that tendencies are not centered (cf. p. [307] in my paper, and Sobel, section 6.42) He wants to stay *neutral* on this issue, while you are prepared to *commit* yourself. This policy of neutrality means that Sobel's theory of tendencies must be so weak as to allow of different possible extensions. One such extension is a theory in which tendencies are centered without being (in general) fully determinate.

[…][22]

Is it, however, reasonable to allow for the possibility that tendencies might be centered (even in indeterministic worlds)? As you point out, in our ordinary talk of chance, we treat chances (tendencies) as non-centered. Thus, for instance, the mere fact that a certain coin falls heads is not supposed to show that the chance of heads was one. I think, however, that this argument from ordinary parlance is not conclusive. In particular, when an act-utilitarian defines the utility of an action, *A*, as the weighted sum of the values of its possible consequences, with weights being the chances of different consequences given the performance of *A*, then his concept of chance might very well differ from the ordinary one. It is often said that what is characteristic for the act-utilitarian approach is its *realism* – its tendency to take into consideration 'all the facts'. But then it is natural for such a realistic approach to take into consideration the actual future outcomes of indeterministic processes. And this seems to imply that the *practical* concept of chance – the one used in the calculation of utility – should be a centered one (even though our ordinary concept of chance is not centered at all).

An example: If an agent who is offered a bet on heads with odds five to one accepts the bet and loses, then my act-utilitarian would say that the agent acted wrongly in accepting the bet. He acted wrongly *because he lost*. And accepting the bet would still be wrong even if the odds offered were even better. Now, if

(a) an action is wrong iff it has lower utility than some of its alternatives,

---

[21] Here, I omit a passage dealing with Lewis' comments on Theorem 4 from Rabinowicz (1982a). As mentioned before, I have decided not to discuss that theorem here. In my reply to Lewis on this matter, I suggest that what Lewis takes to be absurd – that chances could be chancy, as he puts it – will be something to be expected if we take chances to be centered.

[22] Here, I omit a lengthy technical passage in my letter.

and

(a) the utility of an action, *A*, is the weighted sum of the values of its possible consequences, with weights being the 'practical' chances of different consequences given the performance of *A*,

then the claim that accepting the bet was wrong, independently of the odds offered, must imply that the *practical* chance of loss was equal to one. This is so even if we assume that the coin was fair and indeterministic, so that, in ordinary parlance, the chance of loss was only 50%. Thus, the act-utilitarian realism leads to the conclusion that the chances used in the calculation of utility should be centered.

How does this relate to decision theory? Well, insofar as decision theory prescribes maximization of *expected* utility and the expected utility of an action is taken to be the weighted sum of its possible utilities, with weights being the probabilities (credences) of these different utilities, the conclusion is obvious: The concept of chance used in the calculation of expected utility should also be a centered one!

Note that, if practical chances are centered, then it is no longer implausible to define the truth conditions for (centered) subjunctive conditionals in terms of chances, as Sobel does.[23] We must only remember that chances used in the definition should be 'practical', and not the ordinary ones.

If we ignore 'abnormal' cases, we may define ordinary chances in terms of conditional credences and (full) dependency hypotheses – just as you suggest. But then we still have the problem of defining practical chances.

Another thing: Even though practical chances differ from the ordinary ones, we should still expect that, in all 'normal' cases, probable ordinary chances (i.e. credence-weighted sums of possible ordinary chances) will still coincide with probable practical chances.[24]

The above distinction between practical and ordinary chances and the suggestion that practical chances should be centered are due to Sobel. In a letter written in February this year [1982], he writes:

> Weak centering is right for the theory of chance: we want room for agents whose chance-views are weakly centered. But, roughly speaking, strong centering is right for the theory of practical relevance for which we want not 'probable chances' but 'probable centered-chances'.

It is such 'probable centered-chances' that Sobel nowadays wants to use as weights in calculation of expected utility. He develops this suggestion in more detail in a manuscript called 'Notes on some recent changes in the theory of chance and the theory of probable chance' (21 Febr. 1982). Thus, one might say that Sobel no longer wants to remain neutral. He is prepared to commit himself to centering (in practical contexts). As for myself, I feel that this is the right thing to do.

Sobel thinks that 'centered-chances' are easily definable in terms of ordinary non-centered ones. He simply lets the centered-chance of *q* given *p* be equal to the truth-value of *q* (one or zero) *if p* is true, and to the ordinary chance of *q* given *p if p* is false. It seems to me that this definition is wrong. I think that the centered-chance of *q* given *p* might deviate from its ordinary chance even in some cases when *p* is false. In particular, this happens when *q* describes a chance event whose occurrence does not in any way depend on whether *p* is true or false.

Thus, to take the simplest case, if the agent declines an offered bet on heads and the coin subsequently is tossed and falls heads, then – insofar as the outcome of the toss is assumed to

---

[23] As we remember, according to Sobel, the subjunctive conditional "If it were that *X*, then it would be that *Y*" is true in *v* iff $T_{v, X}(Y) = 1$.

[24] I.e., more precisely, in all 'normal' cases, the subjective expectation of the centered chance will be equal to the subjective expectation of the ordinary chance.

be wholly independent of the agent's betting behaviour – we should let the centered-chance of heads conditional on the agent's accepting the bet to be one (even though the agent has, in fact, declined the bet, and the ordinary chance of heads was less than one.)

As a matter of fact (this was something that I was not aware of when I wrote to Sobel[25]), there may be cases in which the centered-chance of *q* given *p* is neither one nor zero but it still differs from the ordinary chance of *q* given *p*.

An example: Suppose that we have two coins, c_1 and c_2, with different ordinary chances for heads. Let these chances be 1/2 and 1/3, respectively. The agent is offered the following bet: He will win iff both coins will fall heads or both will fall tails. We assume that c_1 is going to be tossed whatever the agent does – whether he accepts the bet or not. c_2, on the other hand, will be tossed only if the agent accepts the bet. The ordinary chance of winning conditional on accepting the bet is in this case equal to 1/2 (this chance equals $1/2\times1/3 + 1/2\times2/3$). Suppose, however, that the agent declines the bet and c_1 is tossed and falls heads. (c_2 is not tossed, since the agent has abstained from betting.) Then it seems that the centered-chance of winning conditional on accepting the bet should be taken to be equal to 1/3. (If the agent had accepted the bet, he would have won iff c_2 had fallen heads. Therefore, his centered-chance of winning would equal the ordinary chance of c_2 falling heads.) Thus the centered-chance of winning is in this case neither zero nor one, but it still differs from the ordinary chance of winning.

If I am right, then, defining centered-chances in terms of ordinary chances might prove to be quite complicated.[26]

Concerning abnormal cases, I found your armour example fascinating. Does the information that the agent will survive the coming battle give him any reason to abstain from buying the armour? My first intuitive reaction was negative. It seemed to me that this additional information would not change anything in his decision problem: Buying the armour would still be the reasonable thing to do. All this providing

(a) that survival is much more important for the agent than not becoming a pauper,

and

(b) that buying the armour significantly raises his chance to survive.

For example, suppose that the values of survival and of becoming a pauper are, respectively, 100 and -10,[27] while the probable (ordinary) chances of survival are, respectively, 2/3 and 1/3, depending on whether the agent will buy the armour or not.

But then I started to wonder. After learning that he is going to survive, the agent becomes certain that the action he actually is going to perform will have good consequences. And he knows that the chances are that he might not survive if he were to act differently. Thus, it seems that everything depends on the agent's credences for his two options. In particular, if he believes that he won't buy the armour, then, perhaps, not buying the armour is the rational thing for him to do. On the other hand, if the agent, before learning that he was going to survive, thought it quite credible that he would buy the armour, then the conditionalization model for belief-change implies that buying the armour should become even more credible

---

[25] Unfortunately, the letter to which I refer above is not preserved.

[26] But see the concluding section for a suggestion how such a definition could be constructed.

[27] I.e. the values of different possible outcomes are as follows:

Survival without becoming a pauper:	100
Survival as a pauper:	90
Death without becoming a pauper:	0
Death as a pauper:	-10

The last value on the list might be thought to be somewhat implausible: Becoming a pauper should not matter much if I will die in the battle. But we can assume that I have a family to care for, who will be impoverished.

given the new information. (I assume here that the agent's *initial* credences for survival conditional on his different options go by chances. That is, I assume that

$$C_0(S/A) = 2/3 \text{ and } C_0(S/\neg A) = 1/3.$$

C_0 represents the agent's initial credence function, '*S*' stands for 'survival' and '*A*' denotes 'buying the armour'.

Now, let C_S stand for the agent's credence function after receiving the information that he is going to survive. The conditionalization model implies that

$$\begin{aligned} C_S(A) = C_0(A/S) &= [C_0(S/A)\times C_0(A)]/[C_0(S/A)\times C_0(A) \ + \ C_0(S/\neg A)\times C_0(\neg A)] \\ &= 2/3C_0(A)/[2/3C_0(A) + 1/3 - 1/3C_0(A)] \\ &= 2C_0(A)/[C_0(A) + 1] \end{aligned}$$

Thus, if $0 < C_0(A) < 1$, the ratio $C_S(A)/C_0(A)$ will be higher than one.[28]) But then the new information should not give the agent any reason to abstain from buying the armour. (In fact, the opposite is true.[29])

The centered extension of Sobel's theory bears out the informal argument presented above. Assuming that the values of survival and of becoming a pauper are 100 and -10, and that the ordinary chances of survival given *A* and $\neg A$ are, respectively, 2/3 and 1/3,[30] the centered approach to expected utility implies that

$$U(A) = C_S(A)(100 - 10) + C_S(\neg A)(2/3\times 100 - 10),^{31}$$

while

$$U(\neg A) = C_S(A)(1/3\times 100) + C_S(\neg A)100.^{32}$$

But then it follows that

$$U(A) > U(\neg A) \text{ iff } C_S(A) > 13/30.^{33}$$

Thus, assuming that the agent's initial credences go by chances,[34] and supposing that he changes his beliefs by conditionalization,

---

[28] Explanation: If $C_S(A) = 2C_0(A)/[C_0(A) + 1]$, then, if $0 < C_0(A)$, the ratio $C_S(A)/C_0(A) = 2/[C_0(A) + 1]$. And $2/[C_0(A) + 1] > 1$ if $C_0(A) < 1$.

[29] Explanation: That the ratio $C_S(A)/C_0(A) > 1$, means that my credence in *A* increases upon receiving information *S*. Now, this increase in my credence for *A* should give me even more reason to perform *A*, since the new information at the same time implies that the action I will actually perform is going to lead to a good outcome.

[30] More precisely, it is here assumed that the ordinary chances of *S* given *A* and given $\neg A$ are, respectively, 2/3 and 1/3, in every world *v* such that $C(v) > 0$.

[31] Explanation: $U(A)$ is calculated given the agent's information that *S* holds (i.e. that the agent will survive). We take it that the agent is certain that the ordinary chances of *S* given *A* and given $\neg A$ are 2/3 and 1/3, respectively. Then $U(A)$ equals a weighted sum of *A*'s expected utilities given *A* and given $\neg A$, with weights being the credences of these alternatives, $C_S(A)$ and $C_S(\neg A)$, respectively. Now, on the assumption that *A* will be performed, the centered chance of *S* if *A* were performed equals 1. Therefore, on this assumption, the expected utility of *A* is $1\times 100 - 10 = 100 - 10$ (where 100 is the value of survival and -10 is the disvalue of becoming a pauper). On the other hand, on the assumption that $\neg A$ is the case, the centered chance of *S* if *A* were performed equals the ordinary chance of *S* if *A* were performed, which is 2/3. Therefore, on this assumption, the expected utility of *A* is $2/3\times 100 - 10$.

[32] Explanation: $U(\neg A)$ equals a weighted sum of $\neg A$'s expected utilities given *A* and given $\neg A$, with weights being the credences of these alternatives, $C_S(A)$ and $C_S(\neg A)$, respectively. On the assumption that *A* will be performed, the centered chance of *S* if $\neg A$ were performed equals the ordinary chance of *S* if $\neg A$ were performed, which is 1/3. Therefore, on this assumption, the expected utility of $\neg A$ is $1/3\times 100$ (where 100 is the value of survival). On the other hand, on the assumption that $\neg A$, the centered chance of *S* if $\neg A$ were performed = 1. Therefore, on this assumption, the expected utility of $\neg A$ is $1\times 100 = 100$.

[33] The calculation is left to the reader.

$$U(A) > U(\neg A) \text{ iff } C_0(A) > 13/47.$$[35]

(Other variants of decision theory would lead to different results. Thus, according to the standard, Jeffrey-like approach [= evidentiary decision theory], the agent should abstain from buying the armour – quite independently of his credence for that option.[36] Your own theory would have given the same answer, *if* it were applicable to the abnormal cases.[37] But, of course, it is not.[38] And the non-centered extension of Sobel's theory would imply that buying the armour is always best policy – quite independently of the agent's credence for that option.[39])

Of course, it seems very strange to say that, in some cases, the rationality of an action might be dependent on its credence. But then the abnormal cases, in which we have this kind of dependence, *are* strange. (However, I do not mean to suggest that this kind of dependence can obtain *only* in the abnormal cases. The story of the man who met death in Damascus, described by Gibbard and Harper in their paper 'Counterfactuals and Two kinds of Expected Utility' provides an example of a 'normal' case in which we seem to have the same kind of dependence.[40])

Anyway, the following seems to be true: *If* decision theory should always treat the agent's credences for his options as *irrelevant* for choice, then you are right in suggesting that decision theory is not applicable to abnormal cases.[41]

[…]

---

[34] I.e., assuming that $C_0(S/A) = 2/3$ and $C_0(S/\neg A) = 1/3$.

[35] As we have seen above, if $C_0(S/A) = 2/3$ and $C_0(S/\neg A) = 1/3$, the condionalization model implies that $C_S(A) = 2C_0(A)/[C_0(A) + 1]$. $2C_0(A)/[C_0(A) + 1] > 13/30$ iff $C_0(A) > 13/47$.

[36] If my credence in survival is one, as we have assumed, $V(A)$ – the expected value of buying the armour – is 90. The expected value of not buying the armour, on the other hand, is 100, since the news that I won't buy the armour cannot decrease my credence in my survival, if the latter equals one. Note, however, that this result crucially depends on my being *certain* that I will survive. If my credence in *S* is lower than one, however slightly, the claim made above doesn't hold any longer.

[37] Since $C_S(S) = 1$, $C_S(S/AK)$ must be one, for every dependency hypothesis *K* such that $C_S(K) > 0$. Similarly, $C_S(S/\neg AK)$ must be one, for every such *K*. Consequently, survival is guaranteed whatever choice one will actually make. At the same time, becoming a pauper is guaranteed iff one chooses *A*. This means that for all *K* such that $C_S(K) > 0$, $V(AK) = 100 - 10 = 90$, while $V(\neg AK) = 100$. (For simplicity, I here assume that the only things the agent cares about are survival and not becoming a pauper.) Therefore, on Lewis' formula for expected utility, the expected utility of $A = 90$, while the expected utility of $\neg A = 100$.

[38] As clarified by Lewis in his letter.

[39] The non-centered version of Sobel's theory requires us to calculate the expected utility of an option in terms of the ordinary chances of its effects. This means that the agent must simply ignore the information that he will survive. This information won't have any influence on the expected utilities of his options. On this non-centered approach,

$$U(A) = C_S(A)(2/3\times100 - 10) + C_S(\neg A)(2/3\times100 - 10) = 2/3\times100 - 10,$$

while

$$U(\neg A) = C_S(A)(1/3\times100) + C_S(\neg A)\,(1/3\times100) = 1/3\times100.$$

Since $2/3\times100 - 10 > 1/3\times100$, it follows that $U(A) > U(\neg A)$.

[40] In this famous example, inspired by Somerset Maugham's 'The appointment in Samarra', Death is awaiting the agent in one of two localities, *a* or *b*. The agent, who would like to avoid his appointment with Death, is in a severe quandary: Should he flee to *a* or stay in *b*? We assume that no other option is available. The agent believes that Death is a good predictor of his choices. Therefore, if the agent feels inclined to go to *a*, his credence for that option increases, which in turn increases his credence that Death awaits him in *a*. This gives him a reason to stay in *b* instead, which modifies his inclinations in favour of staying in *b* and thus increases his credence for that option. This increases his credence that Death awaits him in *b*, which gives him a reason to escape to *a*. And so on. The decision problem exhibits a radical instability. (See Gibbard & Harper 1978.)

[41] For an argument that the 'if'- part of the sentence above is questionable, see Rabinowicz (2002). In that paper, the 'abnormal' cases are not discussed, but I use 'Death in Damascus' as an example. For a similar argument, cf. Joyce (2002).

6. Concluding remarks

As should be clear by now, I find the idea of centered chances attractive. The expected utility of an action is the subjective expectation of its utility. Or, to put it more precisely, it is a weighted sum of its utilities in different possible worlds that are compatible with the agent's beliefs, with weights being the agent's credences for these worlds. Since it might well be the case that in some of the worlds with non-zero credence, or perhaps in all of them, the action under consideration is not performed, we should aim at a definition of utility that makes it possible to talk about the utility of an action even in those worlds in which some other action is performed instead. Now, the utility of an action in a world, if understood in this way, depends on what is true in the world in question concerning what would happen if that action were performed. Or at least, but only as the second best, it depends on what could happen under these circumstances. We should first look for 'woulds' and then fill in the remaining gaps with 'coulds', so to speak. Centered chance is the notion that is meant to implement this idea. We need it to spell out the complex counterfactual connection that obtains between an action and its potential outcomes. The utility of an action can be thought of as the sum of the values of its different possible outcomes weighted with the centered chances of these outcomes.

Intuitively, though, centered chance is a notion that should itself be defined in terms of ordinary, non-centered chances. If *X* is true in a given world *v*, the centered chance of *Y* given *X* is one if *Y* obtains in *v* and it is zero otherwise. However, if *X* does *not* obtain in *v*, then the centered chance of *Y* given *X* is determined by the ordinary chance of *Y* conditioned on both *X* *and* all the events that take place in *v* independently of whether *X* obtains in *v* or not. This is the leading intuition, I think, but how are we to make it more precise?

A way to do it is to index ordinary chances with *times* at which they obtain in a given world. That chances can change over time has been argued by Lewis (1980). I believe he is right on this point. If we don't mention the time at which a conditional chance of *Y* given *X* obtains, this is simply because we implicitly think of the time at which *X* would take place if it did. Thus, the time index is implicit in our statement of ordinary chance, but it has to be made explicit if we want to describe changes in chance over time.[42] Let's consider an example: Suppose we have an option of betting on heads in three consecutive coin tosses that are going to take place at t_1, t_2 and t_3. Let's assume that the coin is fair and that coin tossing is an indeterministic process. The bet has to be made, if at all, at t_0. It will be won iff the coin falls heads each time. The coin will be tossed three times whether we accept the bet or not. Now, at t_0, the conditional chance of winning this bet if one were to make it is $1/2\times1/2\times1/2$. Suppose that the coin falls heads in the first toss, at t_1. This means that, at t_1, the conditional chance of winning the bet (if it had been made) increases to $1/2\times1/2$. If, at t_2, the coin again falls heads, the chance of winning the bet increases to ½ at that time. Finally, at t_3, if the coin falls tails in the third toss, the chance of winning the bet (if it had been made) goes down to 0 and then stays at that value forever after.

Given this time-indexed notion of ordinary chance, we can define the *centered* chance of *Y* given *X* as the ordinary chance of *Y* given *X* as it is 'at the end of time', so to speak, when all the issues pertaining to the connection between *X* and *Y* have been resolved. That is, more precisely, the centered chance of *Y* given *X* in world *v* is the limit towards which the time-indexed ordinary chance of *Y* given *X* develops in *v* as time goes to infinity.[43] Letting T^0 stand

[42] For some X's, however, it might be impossible of meaningless to talk about the time of X's occurrence. Then leaving the time-index out is problematic, if chance can change over time. I gloss over this difficulty here.

[43] I am indebted to John Broome (personal communication) for this suggestion of defining ordinary chance 'at the end of time'as the limit of the time-indexed ordinary chance.

for the ordinary chance function and T for the centered function, we thus get the following definition of the latter in terms of the former:

$$T_{v, X} = \lim(T^{0}_{t, v, X}), \text{ as } t \to \infty$$

Normally, of course, the limit for $T^{0}_{t, v, X}(Y)$, for any proposition Y, will be reached in a finite time: The ordinary chance of Y given X might fluctuate, but sooner or later it reaches a value at which it stays at all subsequent times. Thus, in our example with a bet on heads in three consecutive tosses, the limit for the ordinary chance of winning the bet is reached at t_3. This limit turns out to be 0 and so the centered chance of winning the bet is 0 in this example.

As another example, consider the case I discuss in my letter to Lewis. We have two coins, c_1 and c_2, with different ordinary chances for heads: 1/2 and 1/3, respectively. The agent is offered the bet on both coins falling heads or both falling tails. Coin c_1 is going to be tossed whatever the agent does – whether he accepts the bet or not. The other coin will be tossed if the agent accepts the bet, but not otherwise. The time to make the bet is t_0, while the coin tossing is to be done at t_1. At t_0, the ordinary chance of winning conditional on accepting the bet is equal to 1/2 (= 1/2×1/3 + 1/2×2/3). Suppose that, at t_0.the agent declines the bet. Then, at t_1, only one coin, c_1, is tossed. Suppose it falls heads. This means that at t_1 the ordinary chance of winning the bet decreases to 1/3 (which is the chance of c_2 falling heads if it had been tossed) and then remains at that level forever after. Consequently, the centered chance of winning conditional on accepting the bet is equal to 1/3, just as I suggested in my letter.

The definition of centered chance as the limit of ordinary chance unfortunately does not quite manage to express the leading intuition. In particular, the limit definition might in some cases violate this intuitive constraint on centered chances:

In an X-world, the centered chance of Y given X should be 1 if Y is true in that world and it should be 0 otherwise, for all propositions X and Y and all worlds.

Here is a simple example of such a violation, which was suggested to me by Staffan Angere: As is easy to see, the ordinary chance of at least one heads in an infinite sequence of tosses with a fair coin is 1, both at the beginning of the sequence and at each subsequent point. Which means that the limit of the ordinary chance of *at least one heads in that sequence* is 1, as time goes to infinity. Nevertheless, it is *possible* that the sequence in question never results in heads, in which case the centered chance of at least one heads should be 0 (and not 1).

One might think that the way to deal with this problem is to fall back on a two-pronged definition of centered chance: to use the limit definition for the centered chance of Y given X in the *non-X* worlds, and for the X-worlds simply to stipulate that the centered chance of Y given X is 1 if the world under consideration is a Y-world and 0 if it is not. However, this suggestion is unworkable for essentially the same reasons as those that make Sobel's two-pronged definition of centered chance unworkable. (See above, section 5). Thus, to paraphrase my previous counterexample: Consider a proposition Y: *There will be some heads in the infinite sequence of tosses*. Suppose the agent declines to make a prediction as to whether Y is going to be true. (In the original example, the agent declines a *bet* on a proposition under consideration. But, in the present example, Y is not a kind of proposition on which bets can be made.) Suppose, further, that the coin never falls heads in that sequence. Then – assuming that the outcomes of the tosses in the sequence are independent of whether the agent has made any prediction or not – we should let the centered-chance of Y conditional on the agent making a prediction (= proposition X) to be 0, even though the prediction has not been made (X is false) and the limit of the ordinary chance of Y conditional on X is 1. I am sorry to say I don't know how to deal with this objection.

It seems to me, however, that the idea of centered chance as regular chance 'at the end of time' is distinctive enough for this concept to play a useful role in decision theory, even if it

would turn out that a more precise definition of centered chances, which avoids all counter-examples, is unavailable. After all, many important philosophical concepts are like this: We understand them enough to be able to use them in our theorizing, even though we are unable to define them in a sufficiently precise way.

In my view, Lewis should have welcomed this concept of centered chances. As he points out, the ordinary notion of chance is not centered, which makes it impossible to define truth-conditions for his centered counterfactuals in terms of ordinary chances. However, if we proceed as I have suggested above, we can *first* define centered chances T in terms of ordinary time-indexed chances T^0, and only *then* provide truth-conditions for counterfactuals in terms of T, using for this truth-condition the format suggested by Sobel:

$$X \mathrel{\square\!\!\rightarrow} Y \text{ is true at } v \text{ iff } T_{v, X}(Y) = 1.$$

If we wish, we can, of course, first define the triadic similarity relation between worlds in terms of the centered function T and then give the truth conditions for counterfactuals in terms of that similarity relation. The result will be the same.

Another advantage of centered chances becomes clear when we consider the connection between credence and subjective expectation of chance. In general, one would expect that the two should coincide: The agent's conditional credence for Y given X should equal his conditional expectation of the chances of Y being true if X were true, on the assumption that X is true. Thus,

$$C(Y/X) = \Sigma_v C(v/X) T_{v, X}(Y).$$

If chances were interpreted in the ordinary sense,[44] this equality wouldn't hold in abnormal cases, in which the agent takes himself to have a premonition about the outcome of a chancy process. Then, for example, $C(Y/X)$ might be high, even if the ordinary conditional chance of Y given X is low in the X-worlds to which the agent assigns any credence.[45] However, if T is taken to be centered, then abnormal cases will *not* constitute counter-examples to the equality above: If $C(Y/X)$ is high, then the centered chance of Y given X will have to be high in most of the X-worlds that are being assigned non-negligible credence by the agent.

Before I finish, let me consider one other matter. For ordinary conditional chances it is very reasonable to assume the following *invariance condition*: If, in a world v, the chance of Y given X equals k, then that chance is counterfactually independent of whether X is true in v or not.[46] I. e., (i) if X is true, then the chance of Y given X would still have been k if X had been false; and (ii) if X is false, the chance of Y given X would still have been k if X had been true.[47] But this condition does *not* hold for centered conditional chances. Example: Suppose a fair coin is tossed and happens to fall heads. Then the centered chance of it falling heads if tossed is 1. But if the coin hadn't been tossed, then the centered chance of its falling heads if tossed would have been 1/2 instead. This variability of centered chance has consequences for utility, if the latter is defined in terms of centered chances. Utilities of options will in some cases vary depending on whether the options are performed or not. Thus, to use an example from my letter to Lewis, suppose I can bet at odds 5 to 1 on the outcome of a coin toss (i.e., I

---

[44] I.e., if $T_{v, X}$ in the equation above were interpreted as $T^0_{t, v, X}$, where t is the time at which X would take place, if it did take place.

[45] We have seen this in Lewis' example of the agent who considers whether to spend money on armour: The agent's credence for survival is high, even on the assumption he won't buy the armour, but his expected (non-centered) chance of survival without the armour is relatively low.

[46] I am suppressing the time index here, but – as I suggested above - the ordinary chance of Y given X is meant to stand for the chance of Y given X at the time at which X would take place if it did.

[47] In his letter Lewis makes a suggestion along these lines, when he argues that 'chances aren't chancy'.

can pay 1 unit of money for a chance of winning 5 units if the coin toss goes my way), but the coin will be tossed only if I make the bet. Suppose the coin is fair and coin tossing is an indeterministic process, If I bet on heads and the coin falls tails, the utility of my bet is negative, since the centered chance of my winning the bet is 0 under such circumstances. The utility of my bet is therefore equal to -1. But if I had abstained from betting, the centered chance of my winning the bet would have been 1/2, which means that the utility of the bet would have been positive instead: 1/2×(5-1) + 1/2×(-1) = 3/2. Clearly, utility would not be variable in this way if it instead were defined in terms of ordinary conditional chances. The 'non-centered' utility of the bet under consideration is positive, whether I make it or not.

Up to now, we have been interpreting function *V* on the set of possible worlds as the representation of the agent's desires. Suppose we instead take it to be a measure of the objective value of the worlds in question. Thus, on this interpretation, *w* is *better* than *v* if $V(w) > V(v)$. Then the utility of an action *A* in a world *v*, $\Sigma_w T_{v,A}(w)V(w)$, may be seen as the objective utility of that action[48] and we can use this notion of objective utility to distinguish between right and wrong actions, in the standard consequentialist fashion. According to 'objective' consequentialism, an action is *right* iff its objective utility is maximal, as compared with the objective utilities of the alternative actions that stand at the agent's disposal. It is *wrong* iff some alternative has a higher objective utility. If the objective utility of an action can vary depending on whether it is performed or not, then – obviously – the same applies to the action's 'normative status': to its rightness or wrongness. Thus, in our example above, if we assume that winning the bet would have been objectively valuable (and not just valuable for me, the agent), betting on heads was wrong because the bet was lost. But this action would have been right *if* I had abstained: Under these circumstances, the coin wouldn't have been tossed and therefore the objective utility of the bet would have been higher than the objective utility of abstaining.

This means that on the centered approach to chances, objective consequentialism turns out to violate a seemingly very intuitive condition that has been called the *principle of normative invariance*. This condition was originally put forward by me in late 80-ies, but it was first presented in print in Carlson (1995). According to it, the normative status of an action does *not* counterfactually depend on whether that action is performed or not. Thus, if the action the agent performs is right (wrong), then it would have been right (wrong) to perform even if it hadn't been performed. And if the action is not performed by the agent but it is right (wrong) to perform it, then this action would have been right (wrong) to perform even if it were performed.

The main argument for the principle of normative invariance has to do with the nature of practical deliberation. If the normative status of an action sometimes depends on whether the action is done, then the deliberating agent who hasn't yet decided what to do and who wants to decide what to do by asking what ought to be done, is in trouble: Until he knows what he will decide, he is deprived of an information that is relevant for his decision problem: To decide what to do, he wants to know what ought to be done. But he won't know what ought to be done until he knows what he will do, which he won't know until he decides what to do. Thus, it seems that a normative theory that violates normative invariance is not well suited to be a deliberation guide. (Cf Carlson (1995), p. 101. For a more tolerant assessment of such violations, see Howard-Snyder (2008))

Violations of normative invariance are therefore worrying, but perhaps it is something one can get used to live with. We can in any case find some comfort in the observation that the variability phenomena disappear when it comes to *expected* utility. The latter might

[48] As distinct from its expected utility, $\Sigma_v C(v)\Sigma_w T_{v,A}(w)V(w)$, which on this reading is the agent's subjective expectation of the objective utility of the option.

sometimes vary depending on what the agent thinks he will do.[49] But it does not vary depending on what he actually does.

References

Carlson, Erik, 1995, *Consequentialism Reconsidered*, Dordrecht: Kluwer.

Howard-Snyder, Frances, 2008, 'Damned If You Do, Damned If YouDon't!', *Philosophia* 36, pp. 1 – 15.

Gibbard, Allan, and Harper, William, 1978, 'Counterfactuals and Two Kinds of expected Utility', in C. A. Hooker, J. J. Leach and E. P. McClennen, *Foundations and Applications of Decision Theory*, Dordrecht: D. Reidel, pp. 125-62.

Gärdenfors, Peter, 1982, 'Imaging and Conditionalization', *Journal of Philosophy* **79**: 747—760

James M. Joyce, 2002, "Levi on Causal Decision Theory and the Possibility of Predicting One's Own Actions," *Philosophical Studies* 110: 69 – 102.

Lewis, David, 1980, 'A Subjectivist's Guide to Objective Chance', in Richard C. Jeffrey (ed.), *Sudies in Inductive Logic and Probability*, vol. II., Berkeley: University of California Press, 263-93; re-printed with replies to critics in David Lewis, *Philosophical Papers*, vol. II, Oxford: Oxford University Press, 83-132.

Lewis, David, 1981, 'Causal Decision Theory', *Australasian Journal of Philosophy* 59, 5-30; re-printed in David Lewis, *Philosophical Papers*, vol. II, Oxford: Oxford University Press, 305-337.

Lewis, David, 1982, *Letter to Rabinowicz*, 11 March 1982.

Lewis, David, 1983, 'Postscript' to "Causal Decision Theory" – Reply to Rabinowicz', in David Lewis, *Philosophical Papers*, vol. II, Oxford: Oxford University Press, 337-339.

Sobel, J. Howard, 1978 [1980], *Probability, Chance and Choice*, unpublished book manuscript, revised 1980, http://www.utsc.utoronto.ca/~sobel/PrChChcSp80.pdf

Sobel, J. Howard, 1982, Letter to Lewis and Rabinowicz, 17 April 1982.

Rabinowicz, Wlodzimierz, 1982a, 'Two Causal Decision Theories: Lewis vs Sobel', in TomPauli (ed.), *<320311> - Philosophical Essays Dedicated to Lennart Åqvist*, Uppsala: Philosophical Studies no 34, 299 – 321

Rabinowicz, Wlodek, 1982b, *Letter to Lewis*, 5 April 1982.

Rabinowicz, Wlodek, 2002, 'Does Practical Deliberation Crowd Out Self-Prediction?', *Erkenntnis* 57, pp. 91 – 122.

[49] See my discussion in section 5 above.