7

A Representation Theorem for Causal Decision Theory

Having come to grips with the concept of a conditional belief we now return to the problem of proving a representation theorem for causal decision theory. No decision theory is complete until it has been supplemented with a representation theorem that shows how its "global" requirement to maximize expected utility theory will be reflected at the "local" level as constraints on individual beliefs and desires. The main foundational shortcoming of causal decision theory has always been its lack of an adequate representation result. Evidential decision theory can be underwritten by Bolker's theorem and the generalization of it that was established at the end of Chapter 4. This seems to militate strongly in favor of the evidential approach. In this chapter I remove this apparent advantage by proving a Bolker-styled representation result for an abstract *conditional decision theory* whose two primitives are probability under a supposition and preference under a supposition. This theorem is, I believe, the most widely applicable and intuitively satisfying representation result yet attained. We will see that, with proper qualifications, it can be used as a common foundation for both causal decision theory and evidential decision theory. Its existence cements one of the basic theses of this work. It was claimed in Chapter 5 that evidential and causal decision theories should not be seen as offering competing theories of value, but as disagreeing about the epistemic perspective from which actions are to be evaluated. The fact that both theories can be underwritten by the same representation result shows that this is indeed the case.

7.1 THE WORK OF GIBBARD AND ARMENDT

The easiest way to prove a representation theorem for causal decision theory would be to co-opt some existing result by supplementing its axioms with constraints that capture utility as *efficacy value*.

Allan Gibbard has done this by using Savage's theorem, [1] and Brad Armendt has employed Peter Fishburn's conditional utility theory for the same purpose.[2] While both these results are interesting and important for what they tell us about causal decision theory's relationship to two of the standard formulations of expected utility theory, neither is ideal because the representation results of Savage and Fishburn are less than fully satisfactory.

Gibbard supplements the axioms that govern preferences in Savage's theory by two constraints on beliefs about subjunctive conditionals that together suffice to pick out an acceptable state partition relative to which causal utilities may be computed. Expressed in terms of a decision problem $\boldsymbol{D} = (\Omega, \boldsymbol{O}, \boldsymbol{S}, \boldsymbol{A})$, these new axioms are[3]

Definiteness of Outcome: Let X be any proposition in Ω that the agent cares about (in the sense of not being indifferent between X and $\neg X$), and let A be any act in $\boldsymbol{A}$. Then, either $[(A \mathbin{\Box\!\!\rightarrow} X) \,\&\, S] \mathrel{.=.} S$ or $[(A \mathbin{\Box\!\!\rightarrow} \neg X) \,\&\, S] \mathrel{.=.} S$ should hold for any state S in $\boldsymbol{S}$.

Instrumental Act Independence: For any act A and any state S it should be the case that $[(A \mathbin{\Box\!\!\rightarrow} S) \equiv S] \mathrel{.=.} \mathbf{T}$.

The first of these says that the agent must be certain about all the good or evil things that would accompany her acts when any given state obtains. The second says that she must be certain that the states in $\boldsymbol{S}$ are counterfactually independent of what she does. Under these conditions, Gibbard shows, the expected utility that Savage's axioms deliver for any act A will coincide with its efficacy value. This makes it possible for the causal decision theorist to use Savage's representation theorem as a foundation for causal decision theory subject to the proviso that Savage's axioms are only appropriately applied when the two stated conditions hold.

Gibbard's approach does, of course, alter the nature of Savage's theorem since it no longer characterizes expected utility maximization in terms of constraints on preferences alone. Some will see this as a disadvantage, but I think it quite appropriate. As I argued in connection with Bolker's theorem, it is wrong-headed to try to understand prudential reason by reducing the laws of rational belief to the laws of

1 Gibbard (1984).
2 Armendt (1986). Fishburn's theory is developed in Fishburn (1973).
3 Gibbard does not actually use the comparative probability relation to express these requirements. Instead, he assumes a primitive notion of "knowing that" or "being certain" conditional on some proposition being true. So, where he speaks of the agent's *knowing X conditional on C*, I speak of her *being as confident in X* & C *as she is in C*. It should be clear that there is no substantive difference here.

rational desire since this both gives rise to unacceptable forms of pragmatism and forces theorists proving representation results to impose unduly strong structural constraints on preferences. Thus, I have no objection to Gibbard's talk of the agent's beliefs. I do, however, think it is unwise to employ Savage's representation theorem in this context because all the problems associated with it are thereby imported into causal decision theory. The main problems I have in mind are its use of "constant" acts and its inability to handle "small-world" decision making.

Armendt's representation theorem takes Fishburn's *conditional* decision theory as its starting point. The basic concept here is that of the utility of an action A on the hypothesis that some condition E obtains, written here as $\mathrm{U}(A\|E)$. This quantity is governed by the (partition invariant) equation:

Fishburn's Equation. $$\begin{aligned}\mathrm{U}(A\|E) &= \Sigma_S \mathrm{P}(S/E)\mathrm{u}(A\|S)\\ &= \Sigma_j \mathrm{P}(E_j/E)\mathrm{U}(A\|E_j)\end{aligned}$$

where $\{E_1, E_2, \ldots E_n\}$ is any partition of E. For Armendt's purposes the most important thing about this formula is that it allows for a distinction between an act A's *unconditional* utility and its utility conditional on its own performance. These are given, respectively, by

$$\mathrm{U}(A) = \mathrm{U}(A\|\mathbf{T}) = \Sigma_S \mathrm{P}(S)\mathrm{u}(A\|S)$$
$$\mathrm{U}(A\|A) = \Sigma_S \mathrm{P}(S/A)\mathrm{u}(A\|A \,\&\, S).$$

Notice how much $\mathrm{U}(A)$ looks like Savage's Equation, and how much $\mathrm{U}(A\|A)$ looks like Jeffrey's. Indeed, if one had $\mathrm{U}(A\|S) = \mathrm{U}(A\|A \,\&\, S) = \mathrm{U}(A \,\&\, S)$, as one generally does *not*, then the top formula would be Savage's and the bottom one would be Jeffrey's

In a suggestion that bears similarities to Jeffrey's ratificationist proposal, Armendt argues that A's auspiciousness diverges from its efficacy value precisely in cases where unconditional preference for A differs from her preference for A conditional on itself. With respect to the act A_2 of refusing the extra $1,000 in Newcomb's problem, for example, Armendt writes, "it is highly plausible that A_2 [given] A_2 is ranked below A_2, . . . Under the hypothesis that I [refuse], my preference for [refusing] is diminished, since worlds in which I [refuse] are worlds where an empty [bank account] is likely."[4] More generally, he claims that

[4] Armendt (1986, p. 10).

unconditional preference is perturbed by a conditional hypothesis to the extent that the hypothesis carries information which makes a difference to the agent's estimate of [a prospect] P's *value* or *utility*. But since to hypothesize is not to acquire news, preference is not perturbed by alterations in *degree of belief* . . . The hypothesis that a proposition is true or the hypothesis that it is false does not affect the [causal] utility that the agent attaches to it. But sometimes the hypothesis that the actual world is a P-world may carry information about states that the agent (believes are) correlated with P.[5]

Any difference between $U(A\|A)$ and $U(A)$ is, in Armendt's view, an indication that the state partition has been chosen incorrectly and thus that the value of $U(A)$ cannot be confidently used as a guide to action.

To isolate the right state partition (i.e., the one for which $U(A)$ is A's correct causal expected utility) Armendt imposes two additional constraints on the agent's preferences. Skipping over some of the technicalities, his first idea is that the elements of an appropriate state partition will "screen off" differences in value between A and A conditional on itself. In our terms this can be expressed as follows:

Value Screening. For each state S and act A, the decision maker should be indifferent between A given S and A given A & S, so that $A\|S \approx A\|(A \,\&\, S)$.

When this holds, $U(A) = \Sigma_S P(S)u(A\|A \,\&\, S)$ and $U(A\|A) = \Sigma_S P(S/A)u(A\|A \,\&\, S)$. These equations stand to one another as Savage's and Jeffrey's do. As Armendt goes on to observe, the conditional utilities $U(A\|A \,\&\, S)$ can be eliminated in favor of unconditional news values $V(A \,\&\, S)$ as long as there exists at least one partition of "consequences" $\boldsymbol{C}$ such that for any proposition $C \in \boldsymbol{C}$ the agent's utility for A conditional on $(A \,\&\, C \,\&\, S)$ always equals her unconditional utility for $(A \,\&\, C \,\&\, S)$.

Existence of Consequences: There is a partition of propositions $\boldsymbol{C}$ such that $A\|(A \,\&\, C \,\&\, S) \approx (A \,\&\, C \,\&\, S)$ for all states S and acts A and all $C \in \boldsymbol{C}$.

If this is the case then $U(A\|A \,\&\, S) = \Sigma_C P(C/A \,\&\, S)U(A\|A \,\&\, C \,\&\, S)$ and the equation for unconditional utility becomes

$$U(A) = \Sigma_S P(S)[\Sigma_C P(C/A \,\&\, S)U(A \,\&\, C \,\&\, S)]$$

This is a version of the $\boldsymbol{K}$ expectation formula for causal decision theory discussed in Chapter 5. Armendt takes this similarity of

[5] Armendt (1986, p. 10).

form to show that his two constraints capture the appropriate notion of a $\boldsymbol{K}$ partition for use in calculating efficacy values. If this is right, then adding them to the axioms for Fishburn's conditional utility theory will provide a representation theorem for causal decision theory.

This is a nice idea. Armendt is right to think that one should look to a theory of conditional expected utility to find a foundation for causal decision theory, and his characterization of $\boldsymbol{K}$ partitions is quite suggestive. Still, I am not sure that he has made the case for Value Screening as a hallmark of the partitions that should be used to compute causal utilities. It is hard to see how it captures the intuitive notion of an element of $\boldsymbol{K}$ as "a complete description of the ways in which things the agent cares about might depend on what she does," or, indeed, to see where causality comes into the picture at all. How do the indifferences $A\|S \approx A\|(A \ \& \ S)$ and $A\|(A \ \& \ C \ \& \ S) \approx (A \ \& \ C \ \& \ S)$ reflect facts about the causal connection between A and C in the presence of S? I am willing to be open-minded here – it may be that Armendt really has characterized the role that $\boldsymbol{K}$ partitions play in rational preference rankings – but the case needs to be more clearly made.

Even if it is, however, Armendt's result would still not provide a sound foundation for casual decision theory. The problem is not his work, but Fishburn's. While Fishburn's equation is partition-invariant (for a given $\boldsymbol{S}$), and while his representation result does not employ constant acts, it does assume that *mitigators* exist. Recall from Chapter 3 that a mitigator is an act that is able to offset whatever desirable or undesirable things might occur, for example, an act whose performance can make a person indifferent between the prospect of an asteroid destroying all life on earth in the next five minutes and the prospect of peace and prosperity for a millennium. Fishburn's theory includes an axiom which says, in effect, that for any two events E and F there is an action A such that the agent prefers A on the condition that E to A on the condition that F. Let E = "An asteroid hits the earth in the next five minutes and destroys all life" and F = "There is peace, prosperity, and happiness everywhere on earth for a thousand years." Try to think of an appropriate A, but don't try too hard because there is none. As we saw in Chapter 3, representation theorems that make use of constant acts or mitigators are to be avoided because these are the sorts of structure requirements that cannot be explained away by construing them as extendibility conditions. For this reason we cannot use Fishburn's theory as a basis for causal decision theory.

The only live option here, in my view, is the extended version of Bolker's theorem that was proved Chapter 4. As we saw, this result is particularly appealing from the formal point of view. It requires only two nonnecessary structure axioms – a completeness axiom and a non-atomicity condition – both of which are plausible as canons of rationality when viewed as extendibility requirements. Moreover, unlike Fishburn or Savage, Bolker does not make essential use of preferences over prospects that a reasonable agent might regard as impossible, such as constant acts or mitigators. In the next section I will show how to prove a general representation theorem for causal decision theory on the basis of Bolker's theorem. I will do this by proving a general representation result for *conditional* expected utility theory, and then showing how both causal and evidential decision theory can be seen as instances of it.

7.2 A STATEMENT OF THE THEOREM

The representation result we are after assumes an agent facing a decision $\boldsymbol{D} = (\Omega, \boldsymbol{O}, \boldsymbol{S}, \boldsymbol{A})$ whose beliefs are described by a *conditional* likelihood ranking $(\|.>.\|, \|.\geq.\|)$ defined relative to a set of conditions $\boldsymbol{C}$ that contains all the acts in $\boldsymbol{A}$ (and perhaps other propositions), and whose desires are described by a *conditional* preference ranking $(\|\boldsymbol{>}\|, \|\boldsymbol{\geq}\|)$ also defined relative to $\boldsymbol{C}$. Since I will be assuming that both these rankings are complete for the purposes of this proof I will simply use $\|.\geq.\|$ for $(\|.>.\|, \|.\geq.\|)$ and $\|\boldsymbol{\geq}\|$ for $(\|\boldsymbol{>}\|, \|\boldsymbol{\geq}\|)$. This greatly simplifies the presentation, but the reader should remain aware that this completeness requirement should ultimately be dispensed with, so that the full story is told in terms of the incomplete rankings $(\|.>.\|, \|.\geq.\|)$ and $(\|\boldsymbol{>}\|, \|\boldsymbol{\geq}\|)$.

We seek a set of axiomatic constraints on $\|.\geq.\|$ and $\|\boldsymbol{\geq}\|$ that suffice for the existence of a pair of functions $P(\bullet\|\bullet)$ and $V(\bullet\|\bullet)$ defined on $\Omega \times \boldsymbol{C}$ such that

7.1a. $P(\bullet\|\bullet)$ is a supposition function; that is, for each $C \in \boldsymbol{C}$ one has
SUP_1 (*Coherence*): $P(\bullet\|C)$ is a probability on Ω.
SUP_2 (*Certainty*): $P(C\|C) = 1$.
SUP_3 (*Regularity*): $P(X\|C) \geq P(X \,\&\, C)$ when $C \in \boldsymbol{C}$.

7.1b. For each $C \in \boldsymbol{C}$, $V(\bullet\|C)$ gives expected news values computed relative to $P(\bullet\|C)$, so that for all $X \in \Omega$

$$V(X\|C) = \Sigma_W \frac{P(W \,\&\, A\|C)}{P(A\|C)} u(W)$$

where u is a function that assigns an unconditional utility u(W) to each atomic proposition W in Ω.[6]

7.1c. P(•‖•) ordinally represents ‖.≥.‖, and V(•‖•) ordinally represents ‖$\underline{\succ}$‖.

7.1d. P(•‖•) is unique, and V(•‖•) (and hence u) is unique up to the arbitrary choice of a unit and a zero point relative to which utility is measured.

Once we have a theorem like this it is straightforward to use it to underwrite either causal decision theory or evidential decision theory. We merely need to impose constraints on the decision maker's conditional likelihood ranking that are strong enough to determine that it represents her beliefs under the right sorts of suppositions. So, if we are interested in a representation result for evidential decision theory ‖.≥.‖ must satisfy

MOF (comparative version). If $C\,//\,D$.>.$\neg$**T**//D, then $X\,//(C\,\&\,D)$.≥. $Y\,//(C\,\&\,D)$ if and only if $(C\,\&\,X)\,//\,D$.≥. $(C\,\&\,Y)\,//\,D$.

In the presence of the other axioms this uniquely picks out the standard conditional probability P(•/•) as the value of P(•‖•) relative to $\boldsymbol{C}$ = {$C \in \Omega$: C .>. $C\,\&\,\neg C$}, and, more generally, it forces P(•‖•) to be a Réyni–Popper measure.

The causal decision theorist, on the other hand, can impose whatever conditions she thinks necessary to have ‖.>.‖ capture the decision maker's causal beliefs. As we have seen, one natural constraint here is

SUB (Comparative Version). If$(C \rightarrow X)\,\backslash\backslash$**T** .≥. $\neg(D \rightarrow \neg Y)\,\backslash\backslash$**T**, then $X\,\backslash\backslash C$.≥. $Y\,\backslash\backslash D$

Or, when Conditional Excluded Middle holds,

$$(C \rightarrow X)\,\backslash\backslash\mathbf{T} \;.\!\geq\!.\; (D \rightarrow Y)\,\backslash\backslash\mathbf{T} \text{ if and only if } X\,\backslash\backslash C \;.\!\geq\!.\; Y\,\backslash\backslash D$$

Other, stronger requirements could be imposed on the basis of the view that one takes about the proper analysis of causal judgments. Whatever these requirements are, however, they would be always imposed on top of the ones already given and can therefore be neglected in the present context.

Now it might seem as if there is not much being offered to the evidential decision theorist here. They do, after all, already have

[6] Recall that atomic outcomes are act/state conjunctions $A\,\&\,S$ where $A \in \boldsymbol{A}$ and $S \in \boldsymbol{S}$.

Bolker's theorem as the foundation for their theory and this "generalization" really does not do much since "conditional" utilities reduce to unconditional utilities of conjunctions when the supposition function is indicative; that is, $V(X \| C) = V(X \,\&\, C)$ when $P(\bullet \| C) = P(\bullet / C)$. So, the appearance of a "unified" foundation for causal and evidential decision theory might be illusory. I am happy to grant that there is nothing new in this result as it applies to evidential decision theory as standardly formulated. Its advantage is that it allows evidential decision theorists to extend their theory to allow for news values defined in terms of Réyni–Popper functions. There is a good reason for them to want to do so.

As a number of critics have noted,[7] Jeffrey's theory seems to lead to absurd results in cases where an agent is certain about what she will do. When $P(A) = 1$ Jeffrey's Equation sets $V(A) = V(A \vee \neg A)$ and leaves $V(\neg A)$ undefined (though Jeffrey conventionally sets it to 0). The problem with this is that it makes it appear as if "awareness of [one's] preference for [one's] top-ranked option over $A \vee \neg A$ reduces preference to indifference."[8] It thus becomes impossible for one to speak sensibly about the evidentiary value of acts one has irrevocably decided to perform. Jeffrey has responded to this difficulty by (i) distinguishing the utility, V, that represents the agent's desires *before* she makes up her mind from the utility, V_A, that represents her desires after she is sure she will do A, and (ii) pointing out that $V(A \vee \neg A)$ and $V_A(A \vee \neg A)$ may differ.[9] The agent, in other words, need not be portrayed as being indifferent between the "*status quo*" before she becomes certain of A and the *status quo* afterward. With this distinction in place we can say that she sees herself as better off for having done A just in case $V_A(A \vee \neg A) > V(A \vee \neg A)$.

While this is right as far as it goes, it leaves a crucial issue unresolved. On Jeffrey's proposal a person who is certain she will perform A must still assign the same news value to every act incompatible with A, and this makes it impossible for her to compare news values of acts she is sure she will *not* perform. The most she can say is that as far

[7] See, for example, Sphon (1977, p. 113).

[8] Jeffrey (1977, p. 136).

[9] Jeffrey expresses this point by saying that the contradictory proposition need not appear at the same place in the agent's preperformance and postperformance preference rankings. Instead of saying that $V(A \vee \neg A)$ and $V_A(A \vee \neg A)$ may differ, he says that $V(A \vee \neg A) - V(A \,\&\, \neg A)$ and $V_A(A \vee \neg A) - V_A(A \,\&\, \neg A)$ may differ where it is understood that he is keeping the news value of $A \vee \neg A$ set at 0. The more intuitive way of making the point, it seems to me, is to keep the news-value of $A \,\&\, \neg A$ fixed, say at 0, and to let that of $A \vee \neg A$ vary, depending on what the decision maker does. This coheres better with the idea, expressed in the previous chapter, that the goal of action is to produce the best postaction *status quo*.

as auspiciousness goes, *A* is better than the alternatives. This is not an ideal result. Even someone who is sure she will do *A* should still be able to evaluate her other alternatives and say things like “The most auspicious option among the ones I did not choose was *B*” where this not merely a statement about her *past* evaluations of acts, but an expression of her *current* view of the situation. Comparisons like this, after all, often figure into our justifications of acts; for example, I might be sure I am going to do (or did) *A because* I recognize that *A* is better than *B* and that *B* is better than *C*, where *B* and *C* are acts I know I will not (or did not) perform. Jeffrey’s approach lets me say the first thing but not the second. If I am going to be allowed to make discriminations in evidential expected utility among acts I am sure not to perform, then evidential decision theory’s basic equation must be rewritten so that $\mathrm{V}(B)$ can be well defined even when $\mathrm{P}(B) = 0$.

The best way to do this is by substituting a Réyni–Popper measure for the ordinary conditional probability in Jeffrey’s equation so that it becomes $\mathrm{V}(B) = \Sigma_S \mathrm{P}(S//\mathrm{B})\mathrm{u}(S \mathbin{\&} B)$. Since $\mathrm{P}(S//B)$ can be well defined and positive even when $\mathrm{P}(B) = 0$, this allows a decision maker to draw distinctions in news value among actions that she is quite sure she will not perform.[10] A person who has irrevocably decided to take the extra thousand dollars in the Newcomb problem can, for example, still make sense of the idea that refusing it would be a more auspicious act, not just from the perspective of her predecision beliefs but from her *current* epistemic position. A fully adequate account of the auspiciousness of acts will thus need to traffic in Réyni–Popper measures. And, if this is so, then evidential decision theorists are going to need a new representation theorem because Bolker’s only provides an expected utility representation for propositions that are nonnull relative to the decision maker’s preference ranking. This new theorem will need to make sense of true conditional news values that are not mere unconditional news values of conjunctions. The result they will need is the one we are about to prove.

[10] In his unpublished work Frank Doring has independently suggested that Jeffrey’s theory needs to be formulated in terms of Réyni–Popper measures. While Doring is right about this point his main motivation for accepting it has to do with finding a way of making sense of “if only I had done *A*” evaluations of actions. Since these evaluations have a subjunctive character I do not regard them as being appropriately captured by Réyni–Popper measures. To reemphasize, the Réyni–Popper measures are *not* suited to capturing subjunctive beliefs. Thus, a news-value for an act *B* that an agent is sure she will not perform need not be the same as the act’s efficacy value (as the next sentence in the text indicates).

My proof strategy is going to be one of divide and conquer. For each condition C in $\boldsymbol{C}$, define the C-*section* as the unconditional likelihood/preference ranking pair defined by

$$X \mathbin{.{\geq}^{C}.} Y \text{ if and only if } X\|C \mathbin{.{\geq}.} Y\|C$$
$$X \mathbin{\underline{\boldsymbol{>}}^{C}} Y \text{ if and only if } X\|C \mathbin{\underline{\boldsymbol{>}}} Y\|C$$

Since the function defined in 7.1c is a news value for every C, it makes sense to impose the Jeffrey/Bolker axioms of Chapter 4 on each C-section individually. This produces a set of *sectional representations* SR = {(P(•‖C), V(•‖C)): $C \in \boldsymbol{C}$} where each (P(•‖C), V(•‖C)) pair is a Bolker-style representation of its associated C-section. The challenge will be to stitch these sectional representations together in the right way to get a full joint representation for ‖.≥.‖ and ‖$\underline{\boldsymbol{>}}$‖.

It requires three axioms (really two axioms and a general principle) to ensure that the requisite system of sectional representations SR will exist. The first describes the behavior of propositions that the decision maker regards as certainly true.

Axiom$_1$ (*Certainty*). *If $C \mathbin{.{=}.} \boldsymbol{T}$, then $.{\geq}^{C}.$ is identical to $.{\geq}^{\boldsymbol{T}}.$, and $\underline{\boldsymbol{>}}^{C}$ is identical to $\underline{\boldsymbol{>}}^{\boldsymbol{T}}$.*

This says that the supposition of propositions that an agent regards as certain should not alter her beliefs or desires. When she supposes that some proposition C is true, the agent adopts a new set of beliefs $.{\geq}^{C}.$ that makes C certain and does as little damage as possible to her prior opinions $.{\geq}^{\mathbf{T}}.$. Axiom$_1$ merely says that if she already takes C to be certain, then the new belief system that approximates the old one most closely is the old system itself. The requirement that $\underline{\boldsymbol{>}}^{C}$ should not change is a consequence of the fact that supposition is an *epistemic* operation that affects belief directly and alters desires only through the mediation of beliefs. Supposition, in other words, never changes the decision maker's *basic* desires.

The second rationality requirement for C-sections is

Axiom$_2$ (*Conditional Certainty*). $C \mathbin{.{=}^{C}.} \boldsymbol{T}$.

This says that belief given C should be genuinely based on the supposition of C's truth. In the presence of the other axioms, this implies that things that happen when C is false are irrelevant to beliefs and desires conditional on C. Thus, the agent will always regard X as precisely as likely and desirable as X & C when she supposes C.

Our final constraint on C-sections requires beliefs and desires conditional on C to obey the same laws of rationality that apply to unconditional beliefs and desires:

Axiom$_3$ (*Conditional Rationality*). *C-sections must obey the same laws of rationality that apply to unconditional likelihood and preference rankings.*

This demands that $.\geq^C.$ and $\underline{\boldsymbol{>}}^C$ be evaluated with regard to rationality in the same way that any other unconditional likelihood/preference ranking pair would be. It says, in other words, that a person should be bound by the same laws of rationality when she supposes some hypothesis to be true as when she supposes nothing at all. I have been careful to state this principle in a way that does not presuppose any specific analysis of rationality for unconditional beliefs or desires because I believe that its validity should be affirmed independently of any disagreements there may be about the particulars of such an analysis. This is a view with which I think most decision theorists would agree.[11]

That having been said, I mean to defend a version of Axiom$_3$ that does take a stand on the nature of rationality for unconditional beliefs and desires. As I argued at the end of Chapter 5, I think all value is news value. Thus, I will require each C-section to satisfy the axioms employed in the version of Bolker's theorem established at the end of Chapter 4. Accordingly, my official version of Axiom$_3$ will be

Axiom$_3$. *For each $C \in \boldsymbol{C}$, $.\geq^C.$ should satisfy the laws of comparative probability CP_1–$^*CP_8$, $\underline{\boldsymbol{>}}^C$ must obey the Jeffrey/Bolker axioms EDT_1–$^*EDT_9$, and $.\geq^C.$ and $\underline{\boldsymbol{>}}^C$ should jointly obey Coherence.*

In other words, each belief/desire pair $.\geq^C.$ and $\underline{\boldsymbol{>}}^C$ must be *EDT-coherent* in the terminology of Chapter 4. I suspect that at this point even the most conscientious readers will have forgotten what this means. The only two things about EDT-coherence that matter at the moment are that (i) it forces the base algebra Ω to be atomless with respect to $.\geq^C.$, and (ii) it ensures the existence of a joint probability/news-value representation $(P(\bullet\|C), V(\bullet\|C))$ of $.\geq^C.$ and $\underline{\boldsymbol{>}}^C$ in which

[11] See, for example, Savage (1954/1972, p. 78).

P(•‖C) is unique and V(•‖C) is unique up to the arbitrary choice of a zero point and a unit for measuring utility.

I take Axiom$_1$–Axiom$_3$ to be the fundamental laws governing rational belief and desire conditional on a single hypothesis C. They take us some way toward a representation result for conditional likelihood and preference rankings.

Lemma 7.2 (*Existence of Sectional Representations*). *If* $.\geq^C.$ *and* $\underline{\boldsymbol{>}}^C$ *satisfy Axiom$_1$–Axiom$_3$ for every $C \in \boldsymbol{C}$, then there is a sectional representation*

$$\mathrm{SR} = \{(\mathrm{P}(\bullet\|C), \mathrm{V}(\bullet\|C)) : C \in \boldsymbol{C}\}$$

in which

i. $P(\bullet\|C)$ is a countably additive probability on Ω with $P(C\|C) = 1$.
ii. $V(X\|C) = \Sigma_W [P(W \,\&\, X\|C)/P(X\|C)]V(W\|C)$ when $P(X\|C) > 0$.
iii. $P(\bullet\|C)$ represents $.\geq^C..$
iv. $V(\bullet\|C)$ represents $\underline{\boldsymbol{>}}^C$.

Moreover, any other sectional representation for $.\geq^C.$ and $\underline{\boldsymbol{>}}^C$ will have the form

$$\mathrm{SR}^* = \{(\mathrm{P}(\bullet\|C), a_C \mathrm{V}(\bullet\|C) + b_C) : C \in \boldsymbol{C}\}$$

for a_C and b_C real numbers (dependent on C) with $a_C > 0$.

This lemma codifies what we can say about the rationality of a system of conditional beliefs and desires when we restrict our attention to beliefs and desires under the supposition that a single condition is true.

It does not, however, tell us anything about "mixed" beliefs and desires in which a decision maker judges that X is more likely or more desirable given C than Y is given D. Since "mixed" beliefs and desires of this type are important to the evaluation of actions, we need to extend Lemma 7.2 to cover this case. It does not do so automatically. While any full representation for $\|.\geq.\|$ and $\|\underline{\boldsymbol{>}}\|$ is a sectional representation, the converse is not the case. In fact, nothing we have said to this point guarantees that *any* of these sectional representations is a full representation. It is, for example, consistent with Axiom$_1$–Axiom$_3$ that $\|\underline{\boldsymbol{>}}\|$ is intransitive (even though all of its individual sections are transitive). We must introduce additional axioms if we want to establish the existence of a full representation of the desired type.

There has been a great deal of work done on the representation of ordinary conditional probability functions.[12] The result we need, however, is slightly more general than any that can be found in the literature because we want it to be possible for $\|.>.\|$ to describe nonevidential forms of belief revision like imaging. The axiom required to obtain the desired representation is

Axiom$_4$. *For and $C, D, E \in \boldsymbol{C}$ and $W, X, Y, Z \in \Omega$, $\|.\geq.\|$ must satisfy*

- *Normalization: $C\|C .=. D\|D$ and $\neg C\|C .=. \neg D\|D$.*
- *Transitivity: If $X\|C .\geq. Y\|D$ and $Y\|D .\geq. Z\|E$, then $X\|C .\geq. Z\|E$.*
- *Connectedness: Either $X\|C .\geq. Y\|D$ or $X\|C .\leq. Y\|D$.*
- *Dominance: If W and X are logically incompatible, and if Y and Z are also incompatible, then $W\|C .\geq. Y\|D$ and $X\|C .\geq. X\|D$ only if $(W \vee X)\|C .\geq. (Y \vee Z)\|D$.*
- *Regularity: $X\|C .\geq. (X \& C)\|\boldsymbol{T}$.*
- *Solvability: If $X\|C .\geq. Y\|D$, then there exists $X^* \in \Omega$ such that $(X^* \& X)\|C .=. Y\|D$.*

The only one of these that is not self-explanatory (by this point in this book) is Solvability. It is an Archimedean axiom that rules out infinitesimal probabilities. Since $.\geq^C.$ is atomless (by Axiom$_3$), it follows that, for any proposition $X \in \Omega$ that is nonnull with respect to $.\geq^C.$ the set of numbers $\{P(X \& X^*\|C): X^* \in \Omega\}$ will contain every value in the interval from 0 to $P(X\|C)$. Therefore, if $P(X\|C) > P(Y\|D)$ and both these probabilities are real numbers, then $P(X^*\|C) = P(Y\|D)$ should hold for some X^*.

Using Axiom$_4$ one can establish

Theorem 7.3 (*Existence of Probability Representations for Suppositions*). *If $\|.\geq.\|$ and $\|\underline{\boldsymbol{>}}\|$ satisfy Axiom$_1$–Axiom$_3$ and if $SR = \{(P(\bullet\|C), V(\bullet\|C)): C \in \boldsymbol{C}\}$ is any sectional representation for $(.\geq^C., \underline{\boldsymbol{>}}^C)$, then Axiom$_4$ is necessary and sufficient for*

7.1a SUP$_1$: *$P(\bullet\|C)$ is a countably additive probability on Ω.*
SUP$_2$: *$P(C\|C) = 1$.*
SUP$_3$: *$P(X\|C) \geq P(X \& C)$ when $C \in \boldsymbol{C}$.*

7.1c $P(\bullet\|\bullet)$ ordinally represents $\|.>.\|$.

7.1d $P(\bullet\|\bullet)$ is the only function for which 7.1a and 7.1c hold.

[12] See, for example, Fine (1973, p. 29).

(Interested readers can find the proof of Theorem 7.3 in the Appendix of this chapter.) The most important thing to notice about this result is its uniqueness clause, which ensures that every sectional representation for $\|.\geq.\|$ and $\|\underline{>}\|$ must involve *the same* supposition function. This is a consequence of the Ordinal Uniqueness Theorem that was established in Chapter 4. The uniqueness of $\mathrm{P}(\bullet\|\bullet)$ turns out to be crucial in what follows.

To obtain a joint representation for $\|.\geq.\|$ and $\|\underline{>}\|$, we must introduce supplementary axioms to clarify the nature of conditional preference and its relationship to belief. Here are the "desire specific" conditions that pertain to the preference ranking $\|\underline{>}\|$:

Axiom$_5$. *For and $C, D, E \in \boldsymbol{C}$ and $W, X, Y, Z \in \Omega$, $\|.\geq.\|$ must satisfy*

- *Transitivity: If $X\|C \;\underline{>}\; Y\|D$ and $Y\|D \;\underline{>}\; Z\|E$, then $X\|C \;\underline{>}\; Z\|E$.*
- *Connectedness: Either $X\|C \;\underline{>}\; Y\|D$ or $X\|C \;\underline{<}\; Y\|D$.*
- *Invariance of Basic Desires: If W is an atom of Ω, then $W\|C \approx W\|D$ for all $C, D \in \boldsymbol{C}$.*
- *Solvability: Let X and Y be nonnull relative to C, so that $X\|C\ .>.\ \neg C\|C$ and $Y\|C\ .>.\ \neg C\|C$. For any $Z \in \Omega$ and $D \in \boldsymbol{C}$ such that $X\|C \;\underline{>}\; Z\|D \;\underline{>}\; Y\|C$ there exists $X^* \in \Omega$ with $X^*\|C \approx Z\|D$.*
- *Topological Separability: If $\boldsymbol{D}$ is a subset of conditions in $\boldsymbol{C}$ such that, for all $C, D \in \boldsymbol{C}$, one has either $X\|C > Y\|D$ for all $X, Y \in \Omega$ or $X\|C < Y\|D$ for all $X, Y \in \Omega$, then $\boldsymbol{D}$ must be countable.*

The first two principles require conditional preferences to be transitive and connected. The third expresses the idea that supposition does not affect basic desires. The role of the solvability condition here is the same as it was in Axiom$_4$. For a given probability/news value pair $\mathrm{P}(\bullet\|C)$ and $\mathrm{V}(\bullet\|C)$, define the *essential range* of $\mathrm{V}(\bullet\|C)$ as the collection of numbers $\{\mathrm{V}(X\|C): \mathrm{P}(X\|C) > 0\}$. When $\|.\geq.\|$ and $\|\underline{>}\|$ satisfy Axiom$_3$ the essential range of any utility $\mathrm{V}(\bullet\|C)$ will be an interval on the real line. (See Fact 4 in the Appendix at the end of this chapter.) Hence, if $X\|C \;\underline{>}\; Z\|D \;\underline{>}\; Y\|C$ holds when X and Y are nonnull with respect to $\|.\geq.\|$, then in any real-valued representation $\mathrm{V}(\bullet\|\bullet)$ of $\|\underline{>}\|$ it must be the case that $\mathrm{V}(X^*\|C) = \mathrm{V}(Z\|D)$ for some X^*. If this were not so, $\mathrm{V}(Z\|D)$ could not be any real number. The separability condition ensures that the representing function $\mathrm{V}(\bullet\|\bullet)$ can fit into the real line. To get a sense of its meaning, note that if its antecedent holds and if $\mathrm{P}(\bullet\|\bullet)$ and $\mathrm{V}(\bullet\|\bullet)$ are any representations of $\|.\geq.\|$ and $\|\underline{>}\|$, then the essential ranges of all the functions $\mathrm{V}(\bullet\|D)$, $D \in \boldsymbol{D}$, will form a family of *disjoint* intervals on the real line each of which has a nonempty interior. Since there can be at most countably many such intervals, the set $\boldsymbol{D}$ must be countable.

It will be convenient to introduce some terminology in connection with these last two requirements that will set up our next axiom. Say that the sections $(.\geq^{C}.,\ \underline{\boldsymbol{>}}^{C})$ and $(.\geq^{D}.,\ \underline{\boldsymbol{>}}^{D})$ are *linked* when there are nonnull pairs $X_1\|C$, $X_2\|C$, and $Y\|D$ such that $X_1\|C \boldsymbol{>} Y\|D \boldsymbol{>} X_2\|C$. Given Solvability and Separability, this means that, in any representation of $(\|.\geq.\|, \|\underline{\boldsymbol{>}}\|)$, the essential ranges of $\mathrm{V}(\bullet\|C)$ and $\mathrm{V}(\bullet\|D)$ will have a nonempty intersection that contains an open interval of numbers. A *chain* is a countable sequence of sections $(.\geq^{C1}.,\ \underline{\boldsymbol{>}}^{C1})$, $(.\geq^{C2}.,\ \underline{\boldsymbol{>}}^{C2})$, $(.\geq^{C3}.,\ \underline{\boldsymbol{>}}^{C3})$, . . . , such that $(.\geq^{C1}.,\ \underline{\boldsymbol{>}}^{C1})$ is linked to $(.\geq^{C2}.,\ \underline{\boldsymbol{>}}^{C2})$, $(.\geq^{C2}.,\ \underline{\boldsymbol{>}}^{C2})$ is linked to $(.\geq^{C3}.,\ \underline{\boldsymbol{>}}^{C3})$, $(.\geq^{C3}.,\ \underline{\boldsymbol{>}}^{C3})$ is linked to $(.\geq^{C4}.,\ \underline{\boldsymbol{>}}^{C4})$, and so on. Two sections that appear in the same chain are *fettered*.

Since we are aiming for a representation in which desires are represented by real-valued expected utilities defined over a set of atomic propositions it makes sense to demand that any two sections in $(\|.\geq.\|, \|\underline{\boldsymbol{>}}\|)$ should be fettered. This is not strictly required by the existence of the desired representation, but any representation for which it fails will be very, very odd.[13] I propose to rule them out by fiat:

Axiom$_6$. *Any two sections of* $(\|.\geq.\|, \|\underline{\boldsymbol{>}}\|)$ *are fettered.*

It turns out (as a result of Lemma 7.6) that the linking relation is symmetric, so this axiom makes $(\|.\geq.\|, \|\underline{\boldsymbol{>}}\|)$ into one big chain.

Our next axiom is a generalization of the Coherence principle of Chapter 4. It serves as the fundamental principle of rationality connecting $\|.\geq.\|$ and $\|\underline{\boldsymbol{>}}\|$. To introduce it we need yet another a piece of terminology.

Definition. *A test configuration is a four-tuple* $(X_1\|C,\ X_2\|C,\ Y_1\|D,\ Y_2\|D)$ *for which all of the following hold*

- X_1 *and* X_2 *are mutually incompatible propositions such that* $X_1\| > (X_1 \vee X_2)\|C > X_2\|C$.
- Y_1 *and* Y_2 *are mutually incompatible propositions such that* $Y_1\|D > (Y_1 \vee Y_2)\|D > Y_2\|D$.
- $X_1\|C \approx Y_1\|D$ *and* $X_2\|C \approx Y_2\|D$.

[13] For example, there must be real numbers $x > y$ and a proposition X such that (i) the utility $\mathrm{u}(W)$ of every atom W that entails X falls above x, (ii) the utility of every atom that entails and $\neg X$ falls below y, and (iii) there is no proposition Z such that $\mathrm{P}(Z\|X)$ and $\mathrm{P}(Z\|\neg X)$ are both nonzero.

A test configuration gives us a way of comparing X_1's conditional probability given $X_1 \vee X_2$ in the C-section with Y_1's conditional probability given $Y_1 \vee Y_2$ in the D-section. In general, if $(X_1\|C, X_2\|C, Y_1\|D, Y_2\|D)$ is a test condition, and if $\mathrm{P}(\bullet\|\bullet)$ and $\mathrm{V}(\bullet\|\bullet)$ is a representation of $\|.\geq.\|$ and $\|\underline{\succ}\|$, then $(X_1 \vee X_2)\|C \;\underline{\succ}\; (Y_1 \vee Y_2)\|D$ will hold if and only if $\mathrm{P}(X_1\|C)/\mathrm{P}(X_1 \vee X_2\|C) \;\underline{\succ}\; \mathrm{P}(Y_1\|D)/\mathrm{P}(Y_1 \vee Y_2\|D)$ *or, equivalently, if and only if* $\mathrm{P}(X_1\|C)/\mathrm{P}(X_2\|C) \geq \mathrm{P}(Y_1\|D)/\mathrm{P}(Y_2\|D)$. Our next axiom requires the relationships among probabilities determined in this way by the agent's preferences to cohere with her beliefs.

Axiom$_7$ (*Generalized Coherence*). *Let $(X_1\|C, X_2\|C, Y_1\|D, Y_2\|D)$ be a test configuration. Then,*

- *If $X_2\|C \;.=.\; Y_2\|D$, then $(X_1 \vee X_2)\|C \;\underline{\succ}\; (Y_1 \vee Y_2)\|D$ if and only if $X_1\|D \;.\geq.\; Y_1\|D$.*
- *If $X_1\|C \;.=.\; Y_1\|D$, then $(X_1 \vee X_2)\|C \;\underline{\succ}\; (Y_1 \vee Y_2)\|D$ if and only if $X_2\|D \;.\leq.\; Y_2\|D$.*
- *If $(X_1 \vee X_2)\|C \;.=.\; (Y_1 \vee Y_2)\|D$, then $(X_1 \vee X_2)\|C \;\underline{\succ}\; (Y_1 \vee Y_2)\|D$ if and only if $X_1\|D \;.\geq.\; Y_1\|D$.*

All these clauses are different ways of saying that an agent's preferences only force $\mathrm{P}(X_1\|C)/\mathrm{P}(X_2\|C)$ to be greater than $\mathrm{P}(Y_1\|D)/\mathrm{P}(Y_2\|D)$ when her beliefs do not force the opposite inequality to be true.

7.5 CONSTRUCTING THE REPRESENTATION

We now have the resources we need to construct a representation for conditional utility theory. Here is the result:

Theorem 7.4 (*Existence of Conditional Utility Representations*). *If $\|.\geq.\|$ and $\|\underline{\succ}\|$ satisfy Axiom$_1$–Axiom$_7$, then there is a pair of functions $P(\bullet\|\bullet)$ and $V(\bullet\|\bullet)$ defined on $\Omega \times \mathbf{C}$ such that*

7.1a $P(\bullet\|\bullet)$ is a supposition function.

7.1b For each $C \in \mathbf{C}$ and $X \in \Omega$,

$$\mathrm{V}(X\|C) = \Sigma_W \frac{\mathrm{P}(W \,\&\, A\|C)}{\mathrm{P}(A\|C)}\,\mathrm{u}(W)$$

for some function u that assigns an unconditional utility u(W) to each atomic proposition $W \in \Omega$.

7.1c $P(\bullet\|\bullet)$ *represents* $\|.\geq.\|$, *and* $V(\bullet\|\bullet)$ *represents* $\|\underline{\boldsymbol{\geq}}\|$.

7.1d $P(\bullet\|\bullet)$ *is unique, and* $V(\bullet\|\bullet)$ *(and hence u) is unique up to the arbitrary choice of a unit and a zero point relative to which utility is measured.*

I will merely sketch the argument of Theorem 7.4 here, leaving the proofs of the more difficult lemmas to the last section.

To begin, suppose that $(\|.\geq.\|, \|\underline{\boldsymbol{\geq}}\|)$ satisfies the axioms. Lemma 7.2 entails the existence of a sectional representation SR = $\{(\mathrm{P}(\bullet\|C), \mathrm{V}(\bullet\|C)): C \in \boldsymbol{\mathcal{C}}\}$ for $(\|.\geq.\|, \|\underline{\boldsymbol{\geq}}\|)$, and Theorem 7.3 ensures that its suppositional probability $\mathrm{P}(\bullet\|\bullet)$ is unique. It also follows from Lemma 7.2 that any other sectional representation for $(\|.\geq.\|, \|\underline{\boldsymbol{\geq}}\|)$ will have the form SR* = $\{(\mathrm{P}(\bullet\|C), a_C\mathrm{V}(\bullet\|C) + b_C): C \in \boldsymbol{\mathcal{C}}\}$ where $a_C > 0$ and b_C are real constants that depend on C. The key to proving Theorem 7.4 lies in finding the right constants to make SR* a full representation for $(\|.\geq.\|, \|\underline{\boldsymbol{\geq}}\|)$.

We can start by asking how to construct a joint representation for two linked sections $(.\geq^C., \underline{\boldsymbol{\geq}}^C)$ and $(.\geq^D., \underline{\boldsymbol{\geq}}^D)$. To do this we must find a_C, b_C, a_D and b_D, with a_C and a_D positive, such that

7.5. $X\|C \ \underline{\boldsymbol{\geq}}\ Y\|D$ if and only if $a_C\mathrm{V}(X\|C) + b_C \geq a_D\mathrm{V}(Y\|D) + b_D$

whenever $X\|C$ and $Y\|D$ are nonnull. To find these numbers we rely on the following important fact about linked sections:

Lemma 7.6. *If* $(.\geq^C., \underline{\boldsymbol{\geq}}^C)$ *and* $(.\geq^D., \underline{\boldsymbol{\geq}}^D)$ *are linked, then there exists at least one test configuration of the form* $(X_1\|C, X_2\|C, Y_1\|D, Y_2\|D)$.

Since $X_1\|C \approx Y_1\|D$. $X_2\|C \approx Y_2\|D$ holds in any such test configuration we know that the desired values of a_C, b_C, a_D, and b_D must be such that $a_C\mathrm{V}(X_1\|C) + b_C = a_D\mathrm{V}(Y_1\|D) + b_D$ and $a_C\mathrm{V}(X_2\|C) + b_C = a_D\mathrm{V}(Y_2\|D) + b_D$. This forces it to be the case that

7.7. $a_D = a_C[\mathrm{V}(X_1\|C) - \mathrm{V}(X_2\|C)]/[\mathrm{V}(Y_1\|D) - \mathrm{V}(Y_2\|D)]$
$b_D = a_C V(X_1\|C) - a_D V(Y_1\|D) + b_C$

These turn out to be the crucial relationships involved in obtaining a joint representation for a pair of linked sections. Their importance is due to

Lemma 7.8. *If* $(.\geq^C., \underline{\boldsymbol{\geq}}^C)$ *and* $(.\geq^D., \underline{\boldsymbol{\geq}}^D)$ *are linked, and if* $a_C > 0$ *and* b_C *are chosen arbitrarily and* a_D *and* b_D *are defined as in 7.7, then for all propositions* $X, Y \in \Omega$ *such that neither* $X\|C$ *nor* $Y\|D$ *is null one has*

7.5. $X\|C \ \underline{\boldsymbol{\geq}}\ Y\|D$ *if and only if* $a_C V(X\|C) + b_C \geq a_D V(Y\|D) + b_D$

Moreover, the values of a_D and b_D given by 7.7 are the only ones for which 7.5 holds.

Lemma 7.8 is the heart of my representation theorem. Its proof can be found at the end of this chapter.

The method we used to jointly represent $(.\geq^{C}., \boldsymbol{\geq}^{C})$ and $(.\geq^{D}., \boldsymbol{\geq}^{D})$ can be extended to a chain $(.\geq^{C1}., \boldsymbol{\geq}^{C1})$, $(.\geq^{C2}., \boldsymbol{\geq}^{C2})$, $(.\geq^{C3}., \boldsymbol{\geq}^{C3}), \ldots$ Since $(.\geq^{Cj}., \boldsymbol{\geq}^{Cj})$ and $(.\geq^{Cj+1}., \boldsymbol{\geq}^{Cj+1})$ are linked for each j, Lemma 7.6 gives us a test configuration $(X_j\|C_j,\ X_j^*\|C_j,\ Y_j\|C_{j+1},\ Y_j^*\|C_{j+1})$ for each j. Choose an index k at random, and fix $a_k > 0$ and b_k arbitrarily. Lemma 7.8 then lets us use 7.7 to recursively define a *unique* series of pairs of constants $(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k), (a_{k+1}, b_{k+1}), \ldots$ such that

$$X\|C_j \ \boldsymbol{\geq}\ Y\|C_{j+1} \text{ if and only if } a_j\mathrm{V}(X\|C_j) + b_j \geq a_{j+1}\mathrm{V}(Y\|C_{j+1}) + b_{j+1}$$

holds for all j and all $X, Y \in \Omega$. Let's call the construction that results in the sequence of pairs $(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k), (a_{k+1}, b_{k+1}), \ldots$ the *scaling process*.

The importance of this process in the present context is a result of the following result:

Lemma 7.9: *Let $(.\geq^{C1}., \boldsymbol{\geq}^{C1})$, $(.\geq^{C2}., \boldsymbol{\geq}^{C2})$, $(.\geq^{C3}., \boldsymbol{\geq}^{C3}), \ldots$, be a chain in $(\|.\geq.\|, \|\boldsymbol{\geq}\|)$, and suppose that $(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k), (a_{k+1}, b_{k+1}), \ldots$, is the unique sequence of constants that results from the application of the scaling process when $a_k > 0$ and b_k are chosen arbitrarily. Then, for any indices i and j one has*

7.5.* $X\|C_i \ \boldsymbol{\geq}\ Y\|C_j$ *if and only if* $a_i\, V(X\|C_i) + b_i \geq a_j V^{Dj}(Y\|D_j) + b_j$

Again the proof is presented at the end of this chapter.

Lemma 7.9 and Axiom$_6$ make it possible to represent all the sections in $(\|.\geq.\|, \|\boldsymbol{\geq}\|)$ simultaneously. Axiom$_6$ ensures that $(\|.\geq.\|, \|\boldsymbol{\geq}\|)$ is one big chain. So, if one fixes $a_C > 0$ and b_C for some section $(.\geq^{C}., \boldsymbol{\geq}^{C})$, then the scaling process will produce unique scaling constants a_D and b_D for every other section $(.\geq^{D}., \boldsymbol{\geq}^{D})$. Lemma 7.9 then ensures that the resulting sectional representation $\mathrm{SR}^* = \{(\mathrm{P}(\bullet\|C), a_C\mathrm{V}(\bullet\|C) + b_C): C \in \boldsymbol{C}\}$ is a full representation for $(\|.\geq.\|, \|\boldsymbol{\geq}\|)$.

Thus, the axioms we have set down do perform the job for which they were designed: They capture the notion of conditional evidential expected utility. Since any reasonable decision theory should be expressible in this form, all future work on the foundations of rational choice theory ought to be either attempts to weaken the axioms given here without changing the basic result or attempts to capture special types of conditional evidential expected utility by strengthening the basic axioms presented here.

7.6 APPENDIX: PROOFS

In this section the central theorems and lemmas of this chapter are proved in the order in which they were presented. It is assumed throughout that the pair ($\|.\geq.\|$, $\|\underline{\boldsymbol{>}}\|$) satisfies Axiom_1–Axiom_3 and that SR = {(P(•$\|C$), V(•$\|C$)): $C \in \boldsymbol{C}$} is a *sectional* representation of ($\|.\geq.\|$, $\|\underline{\boldsymbol{>}}\|$) whose existence is guaranteed by Lemma 7.2 (which does not itself require proof since it is a consequence of Theorem 4.3). Since SR is a sectional representation of ($\|.\geq.\|$, $\|\underline{\boldsymbol{>}}\|$) we can always rely on its being the case that $P(X\|C) \geq P(Y\|C)$ iff $X\|C \mathrel{.\geq.} Y\|C$ and that $V(X\|C) \geq V(Y\|C)$ iff $X\|C \mathrel{\underline{\boldsymbol{>}}} Y\|C$. We cannot, however, assume that SR is a full representation, in the sense that $P(X\|C) \geq P(Y\|D)$ iff $X\|C \mathrel{.\geq.} Y\|D$ and $V(X\|C) \geq V(Y\|D)$ iff $X\|C \mathrel{\underline{\boldsymbol{>}}} Y\|D$, since this is what we are trying to prove.

In carrying out the proofs I will make free use of four related facts, all established by Bolker,[14] that pertain to probability/utility pairs that satisfy the Jeffrey/Bolker axioms. All assume a probability/utility representation (P, V) that obeys Jeffrey's Equation on *atomless* algebra Ω.

Fact 1. If $P(X) > 0$ and λ is a real number between 0 and 1, then there is a proposition X^* that entails X and is such that $P(X^*) = \lambda P(X)$ and $V(X) = V(X^*)$.

Fact 2. If $P(X)$ and $P(Y)$ are nonzero and if $V(X) > V(Y)$, then there exist mutually incompatible, nonnull propositions X^* and Y^* that entail X and Y, respectively, and are such that $V(X^*) = V(X) > V(Y^*) = V(Y)$.

Fact 3. If $P(X)$ and $P(Y)$ are nonzero and if $V(X) > V(Z) > V(Y)$, then there exist mutually incompatible, nonnull X^* and Y^* that entail X and Y, respectively, and are such that $V(X^*) = V(X) > V(X^* \vee Y^*) = V(Z) > V(Y^*) = V(Y)$.

Fact 4. V's essential range $I = \{V(X): P(X) > 0\}$ is an interval with nonempty interior on the real line.

Since every (P(•$\|C$), V(•$\|C$)) pair in SR satisfies Jeffrey's Equation and is defined over an atomless algebra (by Axiom_3) these results apply to all of them.

[14] See Bolker (1966, lemma 1.17 and lemma 3.5). The two latter facts are fairly obvious consequences of the two former facts.

Here is the first result we need to prove:

Theorem 7.3 (*Existence of Probability Representations for Suppositions*). *If* $\|.\geq.\|$ *and* $\|\underline{>}\|$ *satisfy Axiom*$_1$*–Axiom*$_3$ *and* $SR = \{(P(\bullet\|C), V(\bullet\|C)): C \in \boldsymbol{C}\}$ *is any sectional representation for* $(.\geq^C., \underline{>}^C)$, *then Axiom*$_4$ *is necessary and sufficient for*

7.1a. SUP$_1$: *$P(\bullet\|C)$ is a countably additive probability on Ω.*
SUP$_2$: *$P(C\|C) = 1$.*
SUP$_3$: *$P(X\|C) \geq P(X \& C)$ when $C \in \boldsymbol{C}$.*

7.1c. *$P(\bullet\|\bullet)$ ordinally represents $\|.>.\|$.*

7.1d. *$P(\bullet\|\bullet)$ is the only function for which 7.1a and 7.1c hold.*

Proof. The necessity of Axiom$_4$ is trivial because each of its conditions is necessary for the existence of a probability representation for $\|.\geq.\|$. So, assume $\|.\geq.\|$ obeys Axiom$_4$. Since each $P(\bullet\|C)$ is a probability for which $P(C\|C) = 1$, and since Axiom$_4$ requires $\|.\geq.\|$ to be regular, there is no question that $P(\bullet\|\bullet)$ is a supposition function. Moreover, since each $.>^C.$ is atomless (by Axiom$_3$) it follows from the Ordinal Uniqueness Theorem of Chapter 4 that $P(\bullet\|C)$ is the *only* representation for $.>^C$, and thus that $P(\bullet\|\bullet)$ must be unique if it is a representation for $\|.\geq.\|$.

To prove 7.1c, let $.\geq^C.$ and $.\geq^D.$ be any two sections of $\|.\geq.\|$. For each integer $n > 0$, Fact 1 guarantees the existence of partitions $\{C(j, n): j = 1, 2, \ldots, n\}$ and $\{D(j, n): j = 1, 2, \ldots, n\}$ such that $P(C(j, n)\|C) = P(D(j, n)\|D) = 1/n$ for all j. Since $P(\bullet\|C)$ and $P(\bullet\|D)$ represent $.>^C.$ and $.>^D.$, this entails that $C(j, n) .=^C. C(k, n)$ and that $D(j, n) .=^D. D(k, n)$ for all j, k, and n. The next step is to show $C(j, n)\|C .=. D(k, n)\|D$ for all cases. If things were otherwise, then (without loss of generality) there would have to be an index n such that

$$\begin{aligned} &C(1,n)\|C .=. C(2,n)\|C .=. \ldots .=. C(n,n)\|C \\ &\quad .>. D(1,n)\|D .=. D(2,n)\|D .=. \ldots .=. D(n,n)\|D \end{aligned}$$

The additivity clause of Axiom$_4$ would then entail that

$$\begin{aligned} &[C(1,n) \vee C(2,n) \vee C(j,n)]\|C \\ &\quad .>. [D(1,n) \vee D(2,n) \vee D(n,n)]\|D. \end{aligned}$$

which contradicts the normality clause of the axiom since the $C(j, n)$'s and $D(j, n)$'s each form a partition. Thus, it must be the case that $C(j, n)\|C .=. D(k, n)\|D$ for all j and k.

Define $C^*(m, n) = [C(1, n) \vee \ldots \vee C(m, n)]$ and $D^*(m, n) = [D(1, n) \vee \ldots \vee D(m, n)]$. The Dominance clause of Axiom$_4$ entails that $C^*(m, n)\|C$.=. $D^*(m, n)\|D$ for all m and n. Moreover, since the additivity law of probability implies that $\mathrm{P}(C^*(m, n)\|C) = \mathrm{P}(D^*(j, n)\|D) = m/n$, and since the rational numbers are dense in the reals, both of the following will hold for any $X, Y \in \Omega$:

- $X\|C$.>. $Y\|C$ iff there are indices m and n with $X\|C$.>. $C^*(m, n)\|C$.>. $Y\|C$.
- $X\|D$.>. $Y\|D$ iff there are indices m and n with $X\|D$.>. $D^*(m, n)\|D$.>. $Y\|D$.

If $X\|C$.>. $Y\|D$ we can use the solvability clause of Axiom$_4$ to find an $X^* \in \Omega$ such that $(X^* \,\&\, X)\|C$.=. $Y\|D$. It then follows that for some indices m and n we have

$$X\|C \;.\!>\!.\; C^*(m,n)\|C = D^*(m,n)\|D \;.\!>\!.\; Y\|D$$

and thus that

$$\mathrm{P}(X\|C) > \mathrm{P}(C^*(m,n)\|C) = m/n = \mathrm{P}(D(j,n)\|D) > \mathrm{P}(Y\|D)$$

Thus, $X\|C$.>. $Y\|D$ implies $\mathrm{P}(X\|C) > \mathrm{P}(Y\|C)$. The converse also holds since there will always be some m and n for which the lower set of relationships holds. This will entail that the upper set of relationships holds as well given that $\mathrm{P}(\bullet\|C)$ represents $.\!\geq^C\!.$ and that $\mathrm{P}(\bullet\|D)$ represents $.\!\geq^D\!.$ This completes the proof of Theorem 7.3. ■

For the next lemma we should remind ourselves that a *test configuration* for $(.\!\geq^C\!., \underline{>}^C)$ and $(.\!\geq^D\!., \underline{>}^D)$ is a four-tuple $(X_1\|C, X_2\|C, Y_1\|D, Y_2\|D)$ where $X_1 \,\&\, X_2$ and $Y_1 \,\&\, Y_2$ are both contradictory, and $X_1\|C \approx Y_1\|D > X_2\|C \approx Y_2\|D$. The result we need to prove is

Lemma 7.6. *If $(.\!\geq^C\!., \underline{>}^C)$ and $(.\!\geq^D\!., \underline{>}^D)$ are linked, then there is at least one test configuration of the form $(X_1\|C, X_2\|C, Y_1\|D, Y_2\|D)$.*

Proof: If $(.\!\geq^C\!., \underline{>}^C)$ and $(.\!\geq^D\!., \underline{>}^D)$ are linked then there are nonnull pairs $F\|C$, $H\|D$, and $F^*\|C$ such that $F\|C > H\|D > F^*\|C$. Since $\underline{>}^D$ cannot be indifferent among all propositions (Axiom$_3$) there must be an H^* in Ω that is not ranked with H by $\underline{>}^D$. Assume without loss of generality that $H\|D > H^*\|D$. The possibilities then are

(i) $F\|C > H\|D > F^*\|C \;\underline{>}\; H^*\|D$

(ii) $H\|C > H\|D > H^*\|D > F^*\|C$

In (i), apply the Solvability clause of Axiom$_5$ twice, once to $\underline{>}^C$ and once to $\underline{>}^D$, to obtain nonnull $Z, Z^* \in \Omega$ with $Z\|C \approx H\|D > F^*\|C$

$\approx Z^*\|D$. In (ii), apply the clause twice to $\succeq^C$ to find nonnull Z, $Z^* \in \Omega$ with $Z\|C \approx H\|D \succ Z^*\|C \approx H^*\|D$. In either event one has $X\|C \approx Y\|D \succ X^*\|C \approx Y^*\|D$ for some nonnull propositions X, Y, X^*, and Y^*.

Since $\mathrm{V}(\bullet\|C)$ represents $\succeq^C$ we can apply Fact 2 with $\mathrm{V}(X\|C) > \mathrm{V}(X^*\|C)$ to find disjoint, nonnull propositions X_1 and X_2 that entail X and X^*, respectively, and are such that $\mathrm{V}(X_1\|C) = \mathrm{V}(Y\|C) > \mathrm{V}(X^*\|C) = \mathrm{V}(F^*\|C)$. Doing the same thing with $\mathrm{V}(\bullet\|D)$ and $\succeq^D$ gives us disjoint, nonnull propositions Y_1 and Y_2 such that $\mathrm{V}(Y_1\|D) = \mathrm{V}(Y\|D) > \mathrm{V}(Y_2\|D) = \mathrm{V}(Y^*\|D)$. Since $\mathrm{V}(\bullet\|C)$ represents $\succeq^C$ and $\mathrm{V}(\bullet\|D)$ represents $\succeq^D$ this entails $X_1\|C \approx Y_1\|D > X_2\|C \approx Y_2\|D$. $(X_1\|C, X_2\|C, Y_1\|D, Y_2\|D)$ is the test configuration we seek. This completes the proof of Lemma 7.6. ■

We now turn to the crucial result:

Lemma 7.8. *If $(.\geq^C., \succeq^C)$ and $(.\geq^D., \succeq^D)$ are linked, and if $a_C > 0$ and b_C are chosen arbitrarily and a_D and b_D are defined by*

7.7. $a_D = a_C[\mathrm{V}(X_1\|C) - \mathrm{V}(X_2\|C)]/[\mathrm{V}(Y_1\|D) - \mathrm{V}(Y_2\|D)]$
$b_D = a_C V(X_1\|C) - a_D V(Y_1\|D) + b_C$

for $(X_1\|C, X_2\|C, Y_1\|D, Y_2\|D)$ any test configuration associated with $(.\geq^C., \succeq^C)$ and $(.\geq^D., \succeq^D)$, then one has

7.5. $X\|C \succeq Y\|D$ *if and only if* $a_C V(X\|C) + b_C \geq a_D V(Y\|D) + b_D$

for all propositions $X, Y \in \Omega$ such that neither $X\|C$ nor $Y\|D$ is null. Moreover, the values of a_D and b_D given by 7.7 are the only ones for which 7.5 holds.

Proof. To simplify things, let $a_C = 1$ and $b_C = 0$. The proof works the same way with any other choice. What we want to show first is that for any nonnull X and Y it must be the case that

7.5 $X\|C \succeq Y\|D$ if and only if $\mathrm{V}(X\|C) \geq a_D\mathrm{V}(Y\|D) + b_D$

where

7.7 $a_D = [\mathrm{V}(X_1\|C) - \mathrm{V}(X_2\|C)]/[\mathrm{V}(Y_1\|D) - \mathrm{V}(Y_2\|D)]$
$b_D = V(X_1\|C) - a_D V(Y_1\|D)$

We can assume without loss of generality that $X\|C \succeq Y\|D$, so the goal will be to establish

7.5a $\mathrm{V}(X\|C) \geq a_D\mathrm{V}(Y\|D) + b_D$

The proof will be broken into cases depending on where $X\|C$ and $Y\|D$ fall in relation to elements of the test configuration.

In working through the cases one must keep in mind that the test configuration is such that $X_1\|C \approx Y_1\|D > X_2\|C \approx Y_2\|D$, and that the constants a_D and b_D have been specifically chosen to ensure that $\mathrm{V}(X_1\|C) = a_D\mathrm{V}(Y_1\|D) + b_D$ and $\mathrm{V}(X_2\|C) = a_D\mathrm{V}(Y_2\|D) + b_D$.

Case 1. (1a) $X\|C \geq X_1\|C \approx Y_1\|D \geq Y\|D$

(1b) $X\|C \geq X_2\|C \approx Y_2\|D \geq Y\|D$

In subcase 1a one has $\mathrm{V}(X\|C) \geq \mathrm{V}(X_1\|C) = a_D\mathrm{V}(Y_1\|D) + b_D \geq a_D\mathrm{V}(Y\|D) + b_D$. Subcase 1b is identical with X_1 and Y_1 replaced by X_2 and Y_2.

Case 2. $X_1\|C \approx Y_1\|D$. $X\|C \geq Y\|D$. $X_2\|C \approx Y_2\|D$.

Here we appeal to Fact 3 twice to find nonnull X_1^*, Y_1^*, X_2^*, $Y_2^* \in \Omega$ that entail X_1, Y_1, X_2, and Y_2, respectively, and are such that

$$X_1^*\|C \approx Y_1^*\|D \approx X_1\|C$$
$$(X_1^* \vee X_2^*)\|C \approx X\|C$$
$$(Y_1^* \vee Y_2^*)\|D \approx Y\|D$$
$$X_2^*\|C \approx Y_2^*\|D \approx X_2\|C.$$

Assume, without loss of generality, that $(X_1^* \vee X_2^*)\|C \mathrel{.{\geq}.} (Y_1^* \vee Y_2^*)\|D$ and set

$$\lambda = \mathrm{P}(Y_1^* \vee Y_2^*\|D)/\mathrm{P}(X_1^* \vee X_2^*\|C)$$

(which is sure to be well defined because X_1^* and X_2^* are nonnull in the C-section). Since $1 \geq \lambda > 0$ we can apply Bolker's Fact 1 to find nonnull X_1^{**}, $X_2^{**} \in \Omega$ that entail X_1^* and X_2^*, respectively, and are such that

$$\mathrm{V}(X_1^{**}\|C) = \mathrm{V}(X_1^*\|C) \text{ and } \mathrm{P}(X_1^{**}\|C) = \lambda\mathrm{P}(X_1^*\|C)$$
$$\mathrm{V}(X_2^{**}\|C) = \mathrm{V}(X_2^*\|C) \text{ and } \mathrm{P}(X_2^{**}\|C) = \lambda\mathrm{P}(X_2^*\|C)$$

This entails that $\mathrm{V}(X\|C) = \mathrm{V}(X_1^{**} \vee X_2^{**}\|C)$ since

$$\begin{aligned}
\mathrm{V}(X\|C) &= \mathrm{V}(X_1^* \vee X_2^*\|C) \\
&= \frac{\mathrm{P}(X_1^*\|C)}{\mathrm{P}(X_1^* \vee X_2^*\|C)}\mathrm{V}(X_1^*\|C) + \frac{\mathrm{P}(X_2^*\|C)}{\mathrm{P}(X_1^* \vee X_2^*\|C)}\mathrm{V}(X_2^*\|C) \\
&= \frac{\lambda\mathrm{P}(X_1^{**}\|C)}{\lambda\mathrm{P}(X_1^{**} \vee X_2^{**}\|C)}\mathrm{V}(X_1^{**}\|C) \\
&\quad + \frac{\lambda\mathrm{P}(X_2^{**}\|C)}{\lambda\mathrm{P}(X_1^{**} \vee X_2^{**}\|C)}\mathrm{V}(X_2^{**}\|C)
\end{aligned}$$

The first identity follows from the indifference $(X_1^* \vee X_2^*)\|C \approx X\|C$ and the fact that $\mathrm{V}(\bullet\|C)$ represents $\succeq^C$. The third is a consequence of the identities $\mathrm{V}(X_1^{**}\|C) = \mathrm{V}(X_1^*\|C)$ and $\mathrm{P}(X_1^{**}\|C) = \lambda\mathrm{P}(X_1^*\|C)$, and the fact that X_1^{**} and X_2^{**} are mutually incompatible. The other two identities hold because $\mathrm{V}(\bullet\|C)$ obeys Jeffrey's Equation.

Since $\mathrm{V}(\bullet\|D)$ represents $\succeq^D$ it is also true that $\mathrm{V}(Y\|D) = \mathrm{V}(Y_1^* \vee Y_2^*\|D)$. So, we can establish the desired inequality 7.5a by proving

7.5b $\mathrm{V}(X_1^{**} \vee X_2^{**}\|C) \geq a_D\mathrm{V}(Y_1^* \vee Y_2^*\|D) + b_D$.

Start by using the identities $\mathrm{V}(X_1^{**}\|C) = \mathrm{V}(X_1\|C)$, $\mathrm{V}(X_2^{**}\|C) = \mathrm{V}(X_2\|C)$, $\mathrm{V}(Y_1^*\|D) = \mathrm{V}(Y_1\|D)$, and $\mathrm{V}(Y_2^*\|D) = \mathrm{V}(Y_2\|D)$ to rewrite 7.7 as

7.7* $a_D = [\mathrm{V}(X_1^{**}\|C) - \mathrm{V}(X_2^{**}\|C)]/[\mathrm{V}(Y_1^*\|D) - \mathrm{V}(Y_2^*\|D)]$
$b_D = V(X_1^{**}\|C) - a_D V(Y_1^*\|D)$

This (and a little algebra) allows us to express 7.5b as

$$\frac{\mathrm{V}(X_1^{**} \vee X_2^{**}\|C) - \mathrm{V}(X_1^{**}\|C)}{\mathrm{V}(X_1^{**}\|C) - \mathrm{V}(X_2^{**}\|C)} \geq \frac{\mathrm{V}(Y_1^* \vee Y_2^*\|C) - \mathrm{V}(Y_1^*\|C)}{\mathrm{V}(Y_1^*\|C) - \mathrm{V}(Y_2^*\|C)}$$

Since $\mathrm{V}(\bullet\|C)$ and $\mathrm{V}(\bullet\|D)$ obey Jeffrey's Equation in conjunction with $\mathrm{P}(\bullet\|C)$ and $\mathrm{P}(\bullet\|D)$, respectively, we can rewrite this inequality in terms of probabilities as

$$\frac{\mathrm{P}(X_2^{**}\|C)}{\mathrm{P}(X_1^{**} \vee X_2^{**}\|C)} \leq \frac{\mathrm{P}(Y_2^*\|D)}{\mathrm{P}(Y_1^* \vee Y_2^*\|D)}$$

(Note the change in the direction of the inequality.)

Since X_1^{**} and X_2^{**} were chosen to make it true that $\mathrm{P}(X_1^{**}\|C) = \lambda\mathrm{P}(X_1^*\|C)$ and $\mathrm{P}(X_2^{**}\|C) = \lambda\mathrm{P}(X_2^*\|C)$ where $\lambda = \mathrm{P}(Y_1^* \vee Y_2^*\|D)/\mathrm{P}(X_1^* \vee X_2^*\|C)$ it follows that $\mathrm{P}(X_1^{**} \vee X_2^{**}\|C) = \lambda\mathrm{P}(X_1^* \vee X_2^*\|C) = \mathrm{P}(Y_1^* \vee Y_2^*\|D)$. This allows us to simplify the inequality we need to establish still further to

$$\mathrm{P}(X_2^{**}\|C) \leq \mathrm{P}(Y_2^*\|D)$$

and thus to

$$X_2^{**}\|C \mathrel{.\leq.} Y_2^*\|D$$

since $\mathrm{P}(\bullet\|\bullet)$ ordinally represents $\|.>\|$.

To see why this last relationship has to hold notice that we have a situation in which

- X_1^{**} and X_2^{**} are nonnull, mutually incompatible propositions
- Y_1^* and Y_2^* are nonnull, mutually incompatible propositions
- $X_1^{**}\|C \approx Y_1^*\|D \succ (X_1^{**} \vee X_2^{**})\|C \succeq (Y_1^* \vee Y_2^*)\|C \succ X_2^{**}\|C \approx Y_2^*\|D$

This shows that $(X_1^{**}\|C, X_2^{**}\|C, Y_1^*\|D, Y_2^*\|D)$ is a test configuration. Moreover, since we know that $(X_1^{**} \vee X_2^{**})\|C .=. (Y_1^* \vee Y_2^*)\|D$ it follows from the last clause of the Generalized Coherence principle, Axiom$_7$, that $X_1^{**}\|C .\geq. Y_1^*\|D$, and this implies that $X_2^{**}\|C .\leq. Y_2^*\|D$. 7.5a is thus established in Case 2.

Case 3. $X\|C \succeq Y\|D \succ X_1\|C \approx Y_1\|D \succ X_2\|C \approx Y_2\|D$.

The proof here is similar to that of Case 2, so I will merely sketch the main ideas. Note first that when $X\|C \succ Y\|D$ we can use the Solvability clause of Axiom$_5$ to find a nonnull X^* such that $X^*\|C \approx Y\|D$. Thus, we can assume without loss of generality that we are dealing with a case in which $X\|C \approx Y\|D$ because if we show that $V(X\|C) = a_D V(Y\|D) + b_D$ in this case then $V(Z\|C) > a_D V(Y\|D) + b_D$ will follow whenever $V(Z\|C) > V(X\|C)$. Our goal, then, is to establish that $V(X\|C) = a_D V(Y\|D) + b_D$ given that $X\|C \approx Y\|D \succ X_1\|C \approx Y_1\|D \succ X_2\|C \approx Y_2\|D$.

Using Fact 3 in essentially the same way as in Case 2 we can find propositions X^*, Y^*, X_2^*, $Y_2^* \in \Omega$ that entail X, Y, X_2, and Y_2, respectively, and are such that

- X^* and X_2^* are mutually incompatible
- Y^* and Y_2^* are mutually incompatible
- $X^*\|C \approx X\|C \succeq Y^*\|D \approx Y\|D \succ (X^* \vee X_2^*)\|C \approx X_1\|C \approx (Y^* \vee Y_2^*)\|D \approx Y_1\|D \succ X_2^*\|C \approx Y_2^*\|D \approx X_2\|C \approx Y_2\|D$.

Again, we can assume, without loss of generality, that $(X^* \vee X_2^*)\|C .\geq. (Y^* \vee Y_2^*)\|D$, set

$$\lambda = P(Y^* \vee Y_2^*\|D)/P(X^* \vee X_2^*\|C)$$

and then use Fact 1 to find nonnull X^{**}, $X_2^{**} \in \Omega$ that entail X^* and X_2^*, respectively, and are such that

$$V(X^{**}\|C) = V(X^*\|C) \text{ and } P(X^{**}\|C) = \lambda P(X^*\|C)$$
$$V(X_2^{**}\|C) = V(X_2^*\|C) \text{ and } P(X_2^{**}\|C) = \lambda P(X_2^*\|C)$$

Reasoning analogous to that used in Case 2 yields that $V(X_1\|C) = V(X^{**} \vee X_2^{**}\|C)$, and it follows that $(X^{**}\|C, X_2^{**}\|C, Y^*\|D, Y_2^*\|D)$ is a test configuration in which

$$X^{**}\|C \approx Y^*\|D \succ (X^{**} \vee X_2^*)\|C \approx (Y^* \vee Y_2^*)\|D \succ X_2^{**}\|C \approx Y_2^*\|D.$$

Now, since $V(X^{**}\|C) = V(X\|C)$ and $V(Y^*\|D) = V(Y\|D)$, the relevant version of 7.5a for this case can be established by proving that $V(X^{**}\|C) = a_D V(Y^*\|D) + b_D$. Since $V(X^{**} \vee X_2^*\|C) = V(X_1\|C)$, $V(X_2^{**}\|C) = V(X_2\|C)$, $V(Y^* \vee Y_2^*\|D) = V(Y_1\|D)$, and $V(Y_2^*\|C)$

$= \mathrm{V}(Y_2\|C)$, we can use 7.7 and Jeffrey's Equation to rewrite this inequality as

$$\frac{\mathrm{V}(X^{**}\|C) - \mathrm{V}(X^{**} \vee X_2{}^{**}\|C)}{\mathrm{V}(X^{**} \vee X_2{}^{**}\|C) - \mathrm{V}(X_2{}^{**}\|C)} = \frac{\mathrm{V}(Y^*\|D) - \mathrm{V}(Y^* \vee Y_2{}^*\|D)}{\mathrm{V}(Y^* \vee Y_2{}^*\|D) - \mathrm{V}(Y_2{}^*\|C)}$$

And, in the presence of Jeffrey's Equation, this is equivalent to

$$\mathrm{P}(X^{**}\|C)/\mathrm{P}(X_2{}^{**}\|C) = \mathrm{P}(Y^*\|C)/\mathrm{P}(Y_2{}^*\|C)$$

Given that $\mathrm{P}(X^{**}\|C) + \mathrm{P}(X_2{}^{**}\|C) = \lambda\mathrm{P}(X^* \vee X_2{}^*\|C) = \mathrm{P}(Y^* \vee Y_2{}^*\|D)$, this holds if and only if $\mathrm{P}(X_2{}^{**}\|C) = \mathrm{P}(Y_2{}^*\|C)$ or, equivalently, $Y_2{}^*\|C \mathrel{.{=}.} X_2{}^{**}\|C$ (since $\mathrm{P}(\bullet\|\bullet)$ represents $\|..>\|$). This follows from Axiom$_7$ because $X^{**} \vee X_2{}^*\|C \mathrel{.{=}.} Y^* \vee Y_2{}^*\|D$ and $(X^{**} \vee X_2{}^*)\|C \approx (Y^* \vee Y_2{}^*)\|D$. This completes the proof of 7.5a for Case 3. ■

Case 4. $X_1\|C \approx Y_1\|D > X_2\|C \approx Y_2\|D > X\|C \geq Y\|D$.

This is almost identical to Case 3, and I leave it to the reader.

Since Cases 1–4 exhaust the possibilities, we have shown that 7.5a holds when a_D and b_D are given by 7.7. Showing that no other values do the job is a matter of noting that if both $\mathrm{V}(X_1\|C) = a\mathrm{V}(Y_1\|D) + b$ and $\mathrm{V}(X_2\|C) = a\mathrm{V}(Y_2\|D) + b$ are going to hold, then $b = \mathrm{V}(X_1\|C) - a\mathrm{V}(Y_1\|D) = \mathrm{V}(X_2\|C) - a\mathrm{V}(Y_2\|D)$, and from this it follows directly that $a = [\mathrm{V}(X_1\|C) - \mathrm{V}(X_2\|C)]/[\mathrm{V}(Y_1\|D) - \mathrm{V}(Y_2\|D)]$. This completes the proof of Lemma 7.8.

We turn now to the final lemma we shall need to prove. It assumes a chain of sections $(.{\geq}^{C1}., \geq^{C1})$, $(.{\geq}^{C2}., \geq^{C2})$, $(.{\geq}^{C3}., \geq^{C3})$, ... To simplify the presentation I am going to write these as $(.{\geq}_1., \geq_1)$, $(.{\geq}_2., \geq_2)$, $(.{\geq}_3., \geq_3)$, . . . , and denote their utilities from SR by $\mathrm{V}_j(\bullet) = \mathrm{V}_j(\bullet\,\|C_j)$ for $j = 1, 2, 3, \ldots$.

Lemma 7.9. *Let* $(.{\geq}_1., \geq_1)$, $(.{\geq}_2., \geq_2)$, $(.{\geq}_3., \geq_3)$, . . . *be a chain of sections in* $(\|.{\geq}.\|, \|\geq\|)$, *and suppose that* $(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k), (a_{k+1}, b_{k+1}), \ldots$, *is the unique sequence of constants that results from the application of the scaling process when* $a_k > 0$ *and* b_k *are chosen arbitrarily, so that*

$$a_{j+1} = [\mathrm{V}_j(X_j) - \mathrm{V}_j(X_j{}^*)]/[\mathrm{V}_{j+1}(Y_j) - \mathrm{V}_{j+1}(Y_j{}^*)]$$
$$b_{j+1} = \mathrm{V}_j(X_j) - a_{j+1}\mathrm{V}_{j+1}(Y_j{}^*)$$

where $(X_j\|C_j, X_j{}^*\|C_j, Y_j\|C_{j+1}, Y_j{}^*\|C_{j+1})$ *is a test configuration for each j. Then, for any indices j and k one has*

7.5 $X\|C_j \succeq Y\|C_k$ if and only if $a_jV_j(X) + b_j \geq a_kV_k(Y) + b_k$*

Proof. Assume that $k > j$. The first step is to "thin out" the chain between $(.\geq_j., \succeq_j)$ and $(.\geq_k., \succeq_k)$ so that each section is linked only to its immediate predecessor and successor. There is a simple algorithm for doing this: Go to the least m such that $(.\geq_{m+1}., \succeq_{m+1})$ and $(.\geq_{m-1}., \succeq_{m-1})$ are linked; throw away the section $(.\geq_m., \succeq_m)$; repeat the process until it is no longer possible to do so. This procedure always leaves us with a chain that has $(.\geq_j., \succeq_j)$ as its first link, $(.\geq_k., \succeq_k)$ as its last link, and in which one section is never linked to the section that follows its immediate successor. To keep the proof simple we will simply assume that we had a chain of this sort to begin with.

Fact 4 implies that the essential range $\mathrm{I}_m = \{a_m\mathrm{V}_m(X) + b_m: \mathrm{P}(X\|C_m) > 0\}$ of each utility function $a_m\mathrm{V}_m(X) + b_m$ $(m = j, j + 1, \ldots, k - 1)$ is an interval on the real line with nonempty interior. The previous lemma entails that the sections $(.\geq_m., \succeq_m)$ and $(.\geq_n., \succeq_n)$ are linked if and only if I_m and I_n overlap in an interval with nonempty interior. Thus, since the chain we are dealing with is thin, the intersection of I_m and I_{m+2} must be empty for all $m = j, j + 1, \ldots, k - 1$. The only way for this to happen when each $\mathrm{I}_m \cap \mathrm{I}_{m+1}$ is nonempty is for there to be a real number z_m in each I_m such that either

(I) $x_m \geq z_m \geq y_m$ for all $x_m \in \mathrm{I}_{m-1}$, $y_m \in \mathrm{I}_{m+1}$

or

(II) $x_m \leq z_m \leq y_m$ for all $x_m \in \mathrm{I}_{m-1}$, $y_m \in \mathrm{I}_{m+1}$

The intervals, in other words, must be overlapping and *descending*, as in case (I), or overlapping and *ascending*, as in case (II). Without loss of generality assume that (I) is the relevant possibility. Given that each z_m is in the essential range of $a_m\mathrm{V}_m(\bullet) + b_m$ this means that there must be propositions $Z_j, Z_{j+1}, \ldots, Z_k$ such that

(#) $X\|C_{m-1} \succeq Z_m\|C_m \succeq Y\|C_{m+1}$ for all $X, Y \in \Omega$ and $m = j, \ldots, k - 1$.

And, since we already know from the previous lemma that 7.5* holds for all when $k = j + 1$ we also have

$$(\#\#)\quad a_{m-1}\mathrm{V}_{m-1}(X) + b_{m-1} \geq a_m\mathrm{V}_m(Z_m) + b_{m+1}$$
$$\geq a_{m+1}\mathrm{V}_{m+1}(Y) + b_{m+1} \text{ for all } X, Y \in \Omega \text{ and } m = j, \ldots, k - 1.$$

Now, to complete the proof merely notice that we will either have $k = j + 1$ in which case the chain has only one link and 7.5* follows from the lemma. Or, on the other hand, if $k > j + 1$ then (#) and (##) imply

$$X \| C_j \succeq Z_{j+1} \| C_{j+1} \succeq Y \| C_k \text{ for all } X, Y \in \Omega$$

and

$$a_j V_j(X) + b_j \geq a_{j+1} V_{j+1}(Z_{j+1}) + b_{j+1} \geq a_k V_k(Y) + b_k$$
$$\text{for all } X, Y \in \Omega$$

7.5* follows directly. This completes the proof of Lemma 7.9. ■