

DUHEMIAN THEMES IN EXPECTED UTILITY THEORY*

Philippe Mongin

1. INTRODUCTION

According to the philosophical position of *epistemological holism*, the statements of the empirical sciences do not relate to observations singly, but collectively. This is because these statements belong to logically complex theoretical structures, which are to a large extent indivisible, and also because further theoretical assumptions (an “observational theory”) underlie the observations made to check them empirically. As a consequence of this basic claim, all brands of epistemological holism include an *underdetermination thesis*, to the effect that the scientists’ decisions about hypotheses are underdetermined by the evidence available to them, and in particular, by the results of the tests they perform. Pragmatic reasons must eventually prevail in the choice of attributing the evidence to this or that part of the theoretical whole, and when elaborating on these reasons, philosophers of science will never offer more than partial and context-dependent guidelines.

Duhem’s book, *La théorie physique, son objet, sa structure* (1906, 2nd ed. 1914), henceforth *TP*, stands out as a landmark for the overall position just sketched. The stunning success of this work is largely due to the way in which Quine expanded on it in the celebrated second chapter of *From a Logical Point of View* (1953, II, 6), henceforth *FLPV*. In the brand of epistemological holism favoured by Quine at the time, the set of statements relating to an observation is nothing short of the entire body of accepted knowledge. The accompanying underdetermination thesis claims that the scientists’ response to empirical evidence may involve reconsidering any elements whatever of that large set, including the mathematical or logical ones. Quine goes on to

suggest that the choice between acceptance and rejection is one of mere convenience, and eventually sketches a picture of science as trying to minimize the amount of disturbance brought about by new empirical evidence.

As it turns out, Duhem is much less extreme than Quine in the extent of his epistemological holism. He does not relate physical observations to theoretical statements outside physics, and he typically takes chunks of this discipline to be the collective units of relevance; only exceptionally does he claim that the whole of physics is involved. Consistently, his underdetermination thesis is less dramatic than Quine's, and when he suggests pragmatic remedies, he is both more specific and more convincing than the latter in the wildly speculative conclusion of *FLPV*. Philosophers of science have recognized this difference for some time already, and the post-Quinean catch – the “Duhem-Quine thesis” used to refer to the underdetermination thesis of the two authors at the same time – is not so common today as it once was¹.

To disentangle Duhem and Quine further, we may add that they pursue different strategic aims, and that this disanalogy matters no less than the difference in degree now recognized in the literature. Quine's later work most clearly, but arguably also *FLPV*, gives a *semantic* turn to a position which, for Duhem, was basically *methodological* in character. As is well-known, *FLPV* uses the underdetermination thesis to rebut a dogma of logical positivists – statements can be verified or refuted individually – that underlies their theory of meaning. Although this challenge also matters to methodology, it is not the connection that Quine brings to the fore. Beside the positivist theory of meaning, he wants to debase the further – in his peculiar view, related – dogma that a distinction can be drawn between analytic and synthetic statements, and this is again a semantic target, not a methodological one. Duhem sometimes comes close to discussing meaning issues², but his main purpose is to explain how physicists can make the best out of empirical findings that never warrant indisputable conclusions. We are committed here to a strictly methodological perspective, and will deal only with Duhem's position, leaving aside both Quine's and the irrelevant compound of the early literature. Correspondingly, when we write “holism” in the sequel, we mean epistemological or methodological, not semantic holism. We are aware of the need of clarifying this distinction, but this is not a task for the present paper³.

Following a deep-rooted French tradition running from Bernard to 20th century philosophers like Bachelard and Canguilhem, Duhem developed a local philosophy of

science. He did not mean to extend his conclusions beyond the only science he investigated, which is physics. However, his conclusions can be generalized because they depend on an implicit abstract notion of what a fully grown scientific theory is. The laws of physics, Duhem writes, are both “symbolic” and “approximate”. This can be taken to mean that the laws of physics are formalized, and that at least some of the concepts in the formalized laws can be measured empirically, though always to a degree of approximation. Duhem claim that the laws of physiology have not reached the stage of symbolism and approximation, which leads him to the intriguing conclusion that they evade the problems created by the underdetermination thesis⁴. Using Duhem’s demarcation for what it is worth, we could carry his epistemological holism beyond physics, though not throughout empirical science. The present paper is concerned with applying it to *decision theory*, where some of the generalities counting as putative laws are both “symbolic” and “approximate” in the above sense. Specifically, the paper tries to reinterpret *expected utility theory* in the light of Duhem’s epistemological holism, and it will immediately appear that this particular theory satisfies the two conditions unproblematically. Thus, even on Duhem’s terms, the comparison to come is not unwarranted⁵.

Here is an outline of the overall argument. Section 2 discusses two different *Duhem theses* that can be read in *TP*, i.e., the no-crucial-experiment thesis – crucial experiments, even refuting ones, are impossible – and the underdetermination thesis proper – there is no compelling way of selecting which of the substantial hypotheses, auxiliary hypotheses, and even observation statements, is responsible for an empirical refutation. The distinction of these two theses, which is now widely agreed on, casts further light on Duhem’s holism and facilitates comparison with the falsificationist construals of underdetermination – especially in Popper (1963) and Lakatos (1970) – which we will also address. Most of the paper is concerned with the problem resulting from the underdetermination thesis, i.e., how can an empirical refutation be apportioned across the list of scientific statements so as to licence the rejection of some and the acceptance of others? However, the no-crucial-experiment thesis will also play some rôle in the analysis.

Philosophers of science have been more impressed by the vivid way in which Duhem raises and exemplifies the underdetermination problem than by the few hints he makes to answer it. But the persistent failure of more precise solutions – which we will

document⁶ – has suggested to us that we may fruitfully revisit his suggestions. From *TP* we extracted the following view: although Duhem's underdetermination problem cannot be resolved in a compelling way, it did receive a *satisfactory* solution sometimes in the history of science, and these favourable cases always involved a lengthy process of accumulation and evaluation of conflicting arguments in the face of the evidence, with the arguments typically pointing out in the same direction after a while. Whoever takes this view seriously has no choice but to analyze historical examples painstakingly, as Duhem did in *TP* with his classic account of the theories of light⁷. For the underlying claim is that a satisfactory solution can only be recognized *from the pattern of the arguments proposed over a sufficiently long period of time*, and this imposes on the philosopher of science the task of considering the successive stages of his object-theory, and not just its final or received state. It is this connection precisely – and not just the familiar point that historical examples are useful in clarifying and supporting philosophy of science claims – which has prompted the case study of this paper.

Accordingly, sections 3 and 4 reconstruct part of expected utility theory (EUT) with a view of submitting it to the Duhemian grid. These two sections are mostly technical and historical, with the main assessment awaiting the final section 5. We have skipped many details and actually restricted attention to the simpler of the two branches of EUT – the von Neumann-Morgenstern (VNM) theory of risky choice, which takes probabilities for granted instead of deriving them, as in Savage's more sophisticated contribution. Section 3 starts with *Theory of Games and Economic Behavior* (1944-1947), and section 4 finishes with the experimental work of the 1980-90, when the utility theory originating in this book appeared to be superseded by more promising alternatives. Rather than von Neumann and Morgenstern themselves, we emphasize their immediate followers, like Marschak, who gave their work its final axiomatic touch. Unsurprisingly to those acquainted with the field, the turning point of our narrative is the attack launched in 1952 by Allais against VNM theory and EUT generally, which could then rely on champions of no lesser fame than Marschak, Friedman, Savage and Samuelson. We discuss the major choice experiments that were made in the 1970s and the early 1980s after Allais's suggestions, before we move to some of the burgeoning theoretical developments of the time. This is inevitably very selective, and the interested reader is referred to the excellent surveys available in the field both then and now⁸. The drift of a long story made short is that VNM theory

became swamped with both systematic counterexamples – “effects” in decision-theoretic parlance – and alternative hypotheses that went some way towards accounting for these counterexamples, a situation which naturally suggests trying a Duhemian account of the whole sequence.

Section 5 first discusses the sense in which VNM theory can be said to be refuted by the choice experiments. This conclusion is far from obvious given the wide range of possibilities for empirical refutation that the narrative brings out. There is the further complication that a good deal of the debate, including Allais’s critique, was concerned with the *normative* problem of capturing individual rationality under risk suitably. Despite the obstacles, we will conclude that a Duhemian philosopher could endorse the conclusion in view of the pattern over time of the conflicting arguments. The section also investigates the way in which decision theorists disentangled the possible sources of refutation *within* the theory, i.e., between the axioms, and what rôle exactly was played by surrounding non-VNM hypotheses in the various attempts at providing a solution. Two major results of the inquiry are, for one, that more effort could and should have been done to identify the axiom responsible for the refutation, and for another, that the other hypotheses did not condition the various Duhemian decisions in any important sense. We eventually reject the Lakatosian claim that a genuine refutation requires a theoretical alternative to be available.

As a bibliographical aside, the author may perhaps mention that this is not the first time that he has grappled with Duhemian themes in expected utility theory. Mongin (1988) explored them already, and his paper was actually one the first sustained attempts at discussing holism and underdetermination in the context of – broadly speaking – economics⁹. This early work has since been superseded, both by ensuing changes in the object-theory and a more refined philosophical perception of it. Here, we catch up with these developments, but do not extend the study beyond its initial time limits. The better hindsight on the same forty-year sequence is sufficient to make the account significantly novel.

2. DUHEM’S THESES

2.1 *A statement of Duhem’s underdetermination thesis*

The following passage from *TP* (II, ch. VI, §II, p. 280-281; English trans., p.185) is an elaborate statement of the underdetermination thesis:

A physicist decides to demonstrate the inaccuracy of a proposition; in order to deduce from this proposition the prediction of a phenomenon and institute the experiment which is to show whether this phenomenon is or is not produced, in order to interpret the results of this experiment and establish that the predicted phenomenon is not produced, he does not confine himself to making use of the proposition in question; he makes use also of a whole group of theories accepted by him as beyond dispute. The prediction of the phenomenon, whose non production is to cut off debate, does not derive from the proposition challenged if taken by itself but from the proposition at issue joined to that whole group of theories; if the predicted phenomenon is not produced, not only is the proposition questioned at fault, but so is the whole theoretical scaffolding used by the physicist. The only thing the experiment teaches us is that among the propositions used to predict the phenomenon and to establish whether it would be produced, there is at least one error; but where this error lies is just what it does not tell us. The physicist may declare that this error is contained in exactly the proposition he wishes to refute, but is he sure it is not in another proposition? If he is, he accepts implicitly the accuracy of all the other propositions he has used, and the validity of his conclusion is as great as the validity of his confidence.

A striking feature of this passage is that it describes the experimental physicist as wishing to achieve refutations, not verifications, nor even confirmations – if, as many do, one takes this third notion to be a weakening of the second. Perusal of *TP* suggests that Duhem would accept that much of Popper's philosophy: *empirical evidence informs the scientist via refutations*. Consistently with this stand, Duhem's underdetermination thesis can only mean that an informative piece of evidence undermines a certain conjunction of statements among which it does not help to identify the false ones. This quasi-Popperian reading of *TP* seems to have been accepted nearly universally – albeit tacitly, which comes as a surprise¹⁰.

The selected passage conveys a variety of reasons to conclude that scientific theories are underdetermined by empirical evidence, in the refutationist sense just stated. It argues that physicists must hypothesize a large number of premisses in order to deduce a test statement, i.e., one which can be compared logically with those of an observational record. If this statement clashes with the observational record, one of the premisses must be false, and this is all one can conclude from propositional logic. Within the premisses, there is a natural distinction to be drawn, even if it is not recorded as such by Duhem, between *primary* and *auxiliary* hypotheses. The former belong to the theory under investigation, and typically reappear from one experiment – more generally, one empirical test – to another. Hypotheses of the latter group must be added if the test statement is to be deduced from the former, and they are often required only for the test at hand. But this is not all there is to Duhem's underdetermination thesis.

Although this may seem subdued in the passage, he is not willing to consider the observational record as any more secure than the premisses from which the test statement conflicting with this record is derived. So we should conclude that for Duhem, the set of scientific statements creating the indeterminacy is really *threefold*: (i) primary hypotheses, (ii) auxiliary hypotheses, and (iii) the observational record. Each item can be recognized in the second sentence: “in order to deduce from this proposition the prediction of a phenomenon and institute the experiment which is to show whether this phenomenon is or is not produced”.

Without doing full justice to the complex historical examples in *TP*, this scheme fits with them acceptably well. When discussing Foucault’s celebrated experiment on the speed of light, Duhem only mentions statements contained in group (i), but his account of another experiment in optics, Wiener’s – which was intended to refute an earlier significant hypothesis by Neumann – underscores the relevance of groups (ii) and (iii)¹¹. By analogy with this example, Foucault’s experiment can be analyzed in terms of the three categories. Here, (i) stands for the corpuscular theory of light, (ii) for the various assumptions Foucault had to make in order to measure the velocity of light in water and in the air, and (iii) for the observational record of the respective velocities, once the experiment was performed. Recall that the corpuscular theory entailed the putative regularity that light was *faster* in water than in the air. Foucault’s experimental setting, whereby the speed of light could be measured in both environments, turned this statement into a test prediction, which apparently failed.

The threefold distinction can be refined if necessary. For instance, if one is concerned with explaining rather testing, one may, in the deductive-nomological fashion, subdivide the observational record (iii). There is, for one, a statement of initial conditions (here roughly, “in the experimental setting, light travels through both environments”), and for another, a statement of the experimental result (here roughly, “in the experimental setting, light is *less fast* in water than in the air”). However, if the purpose is only to explore the underdetermination thesis, it is acceptable to collect these statements into a large conjunction, and this is indeed the format of the received discussions of Duhem and Quine¹². Notice further that our classification cuts across logical distinctions made in terms of levels of generality. By nature, (iii) is made out of singular statements, but both (i) and (ii) can contain singular statements, such as those

assigning a value to a physical constant, beside universal statements of either putative laws or *de facto* regularities.

Abstracting from his physical applications, we can state Duhem's underdetermination thesis as follows:

(a) *When a theory is tested against some piece of evidence, and this results in a contradiction between the test prediction and the observational record, there is no compelling way of deciding that some of the primary hypotheses, rather than some of the auxiliary hypotheses or some of the statements of the observational record, are false.*

This roughly corresponds to what has been called “the weak form of the Duhem-Quine thesis”, in contrast with a “strong form”, which in the present framework would read as follows.

(a') *When a theory is tested against some piece of evidence, and this results in a contradiction between the test prediction and the observational record, it is possible to change the auxiliary hypotheses in such a way that they entail the observation record in the presence of unchanged primary hypotheses.*

The so-called strong form cannot be found anywhere in *TP*, but is evocative of a famous claim in *FLPV*: “Any statement whatever can be held true come what may, if we make drastic enough adjustments elsewhere in the system” (1953, p. 43). It is indeed in connection with Quine alone that it has normally been addressed¹³.

These formulations are more or less informative, depending on what is put in the essential words “compelling” in (a) and “possible” in (a'). It is no great news that the truth-tables of propositional logic do not decide where the falsity lies in a false conjunction; so if one actually means “compelling in the sense of standard logic”, there is something painfully obvious about (a). By the same token, “possible” must be restricted so as to preserve (a') from the crude move that consists in replacing the auxiliary assumptions with a statement to the effect that the primary hypotheses imply the observational record. Nevertheless, the two variants do not react similarly when trite possibilities are allowed. For even if one takes “compelling” in the standard logical sense, one learns something useful from (a), which is *the range of statements to which the refutation decision can be applied*. Duhem identifies three possible sources of falsity, but another philosopher, even another holist, might have concluded differently. In other words, what matters is not so much the thesis itself as the kind of

epistemological holism from which it derives. By contrast, nothing remains of (a') if one does not restrict the meaning of "possible" appropriately. As it happens, those philosophers – prominently Grünbaum – who were prepared to give some mileage to the "strong Duhem-Quine thesis" have been unable to clarify what would in this context distinguish an acceptable from an unacceptable change in the auxiliary hypotheses¹⁴. If one were to ask empirical scientists to tell on particular examples what this difference is, counterexamples to (a') would presumably abound. The underdetermination thesis understood this way seems to leave no middle ground between its trivial interpretation and the claim that it is false¹⁵.

Both because of the conceptual difficulties surrounding (a') and the fact that Duhem never considered it, we will take (a) to express the underdetermination thesis. Even this apparently unassuming formulation raises a number of problems, which we will now briefly discuss.

2.2 *On interpreting Duhem's underdetermination thesis*

One problem has already been touched on. If the choice of a logic, *in the sense of some prescriptive formalism*, is left open, (a) makes a very assertive claim. Thus, beyond what can obviously be said for the propositional logic case, it would deny that the Bayesian conditional probability apparatus, the theory of statistical inference, and – for what can be made out of it – the Popperian calculus of "corroboration", can decide which hypothesis is responsible for a refutation. (With a Bayesian or statistical interpretation, "contradiction" would have to be replaced in (a) by a more appropriate word.) An even stronger claim is that there is no *rational* selection of the statements refuted in the conjunction, and formulations like this have indeed circulated in connection with the "Duhem-Quine thesis"¹⁶. However, the specialized literature has usually stuck to the reading of (a) in terms of propositional logic alone, and we think that it is right on this score. For there is a natural division of issues here – first, what does the range of underdetermination, in the straightforward propositional sense, exactly consist of, and second, can this range be narrowed down by moving to some other, more powerful, prescriptive formalism? For clarity, the last question should be treated not under the heading of the underdetermination thesis itself, but as an answer to the selection problem it raises¹⁷.

Second, formulation (a) involves the refutationist analysis of empirical tests, but what about its *verificationist* variant? The suggestion has occasionally been made that the underdetermination thesis should apply to refutation and verification symmetrically. The variant would say that there is no compelling way, in the propositional logic sense, of deciding which hypotheses are verified when the observational record entails the test prediction. Such an underdetermination thesis can only be of interest to those who think that statements can be verified by their consequences – not a large crowd in past and present philosophy of science, and Duhem is not a member of it. The quest for a counterpart to the refutationist variant makes more obvious sense if probabilistic connections replace propositional inferences in the analysis of empirical tests, but this also means that verification gives way to probabilistic confirmation, and an altogether different range of problems arises¹⁸.

Third, there will typically be several primary hypotheses to consider when (a) is applied. Duhem made this point forcefully in connection with Foucault's experiment; in his words, this physicist's concern was the corpuscular *system* ("système de l'émission"), and not the corpuscular *hypothesis* ("hypothèse de l'émission", *TP*, II, ch. VI, §II, p. 283; English trans., p. 187). More will be said on this score after the next, related comment.

Fourth, Duhem's examples also make it clear that auxiliary hypotheses are multifarious; they are not only singular statements or lower-level generalities, as the particular test requires, but also law-like statements that may compare with the primary hypotheses in terms of depth and coverage. Thus, beside highly specific assumptions, Wiener had to assume the standard laws of optics when he set up his experimental device; in this case, the auxiliary hypothesis was no less substantial than Neumann's hypothesis itself. This suggests that the distinction between groups (i) and (ii) should be drawn in *pragmatic* and *epistemic* terms rather than in logical or semantic ones. Roughly speaking, (i) collects those theoretical statements which motivate the empirical inquiry and are currently regarded as problematic, while those in (ii) are introduced on grounds of convenience and at least in the initial intent, not taken to be problematic. Observe that group (i) would be inflated dramatically if it were assumed to include a theory of the scientist's perception. The standard laws of optics, as well as many other physical and non-physical laws, would then be involved quasi-automatically in the range of underdetermination. As it appears from *TP*, Duhem resists this slippery line.

His comments on Wiener relate to the optical theory underlying the physical device, not that underlying the physicist's unaided visual perception. Admittedly, some passages of *TP* extend holism to the point of fusing the sets (i) and (ii), and equating them with the whole of physics, but even there, Duhem does not seem to endorse the strong argument we are now envisaging¹⁹.

Previous discussions of the underdetermination thesis have relied on a separation different from the present one between *one* privileged hypothesis *H*, the various auxiliary hypotheses *A*, and the observational record *O*. Underlying this scheme is the view that an empirical test is directed towards a specified hypothesis. All other hypotheses in (i), even if they belong to the same scientific theory, are pooled with the auxiliary assumptions of (ii). The pool contains all the statements that the scientist can logically blame for the refutation while having no pragmatic or epistemic interest in doing so, since he wishes to relate *O* to *H*, not to any of the *A*. This schematic formalism is sufficient for some argumentative purposes, like the comparative assessment of the “strong” and “weak Duhem-Quine thesis”, but misleading as a rendering of Duhem's thesis in general. It does not fit with Foucault's case, in which a complex theory was subjected to the test. If it were argued that the formalism takes this case into account because *H* can stand for a whole theory, the objection would become that this formalism is too loose, because it hides the difference between Foucault's case and that in which *H* effectively refers to a single hypothesis. There is another significant case to consider, which we will indeed illustrate with EUT: the scientist aims at checking individual components, but in actual practice, tests only large chunks of the theory, because some components turn out to be inseparable. The diversity of empirical situations suggests replacing the standard formalism by the slightly richer representation adopted here.

2.3 Duhem's thesis against crucial experiments

Before Quine advertised Duhem for his epistemological holism, he had been noted for his unusual dismissal of *crucial experiments*. The argument is carried out in a condensed paragraph of *TP* (II, ch. VI, §III) significantly entitled “A ‘crucial experiment’ is impossible in physics” and makes use of the theory of light again. This example is right to the point, because Foucault's experiment was implemented after a

preexisting scheme by Fresnel, who had invented it with the view of finally resolving the conflict between corpuscular and undulatory hypotheses. It would appear that Fresnel conceived of the experiment, if it could be carried out at all, as being both decisive for one hypothesis and against the other. This is the strong sense commonly given to crucial experiments in empiricist philosophies of science, and also the sense that Duhem is concerned with dismissing, but his argument turns out to be relevant to a weaker notion as well. Let us define a *simply refuting* crucial experiment as an experiment which decides against – in the sense of refuting – one of the two theories without necessarily deciding for – in the sense of either verifying or confirming – the other. This restricted meaning is due to Popper, who, from the *Logic of Scientific Discovery* onwards, has repeatedly claimed that it was both important for science and immune to Duhem's criticisms²⁰. With this internal distinction at hand, we state Duhem's second thesis:

(b) *There are no crucial experiments in empirical science, even simply refuting ones.*

When Duhem begins to deny the existence of crucial experiments, he seems only to expand on his already explained underdetermination thesis, but this is deceptive appearance. The two claims are clearly distinct, and it is implausible that Duhem identifies them. In fact, he derives (b) from the same arguments he had used for (a), relying on the continued analysis of the same example. In Fresnel's experimental scheme, a light ray is sent through water, another through the air, one above the other, and a rotating mirror reflects each, producing two spots of light moving on a screen. Supposedly, the location of the spots indicates which of the three following statements holds true: light is as quick in one medium as the other, quicker in water, as the corpuscular hypothesis lets one to expect, or quicker in the air, as the undulatory hypothesis would rather have. Now, the arguments for thesis (a) could be repeated with a similar damaging effect on (b): the predictions on respective speeds follow from whole systems of hypotheses, the postulated connection between these test predictions and the observation of moving spots depends on auxiliary hypotheses, etc. More briefly, if (a) holds, (b) follows, since a crucial experiment is at least decisive against one of the two hypotheses, and (a) says that hypotheses are never decisively refuted. It is important to state the argument in this abstract fashion, which is not Duhem's, because it then becomes clear that it hits the notion of a simply refuting experiment no less than the stronger, more familiar notion.

Duhem has another – by now classic – argument against crucial experiments, which is entirely different from the use of either (a) or the reasons for (a). In a nutshell (and again in abstract terminology), when scientific hypotheses conflict with each other, they are not *contradictory*, but simply *contrary*, with each other; that is, they cannot be true together, but can be false together. *TP* makes this point when denying that two hypotheses alone can ever exhaust the physical possibilities (“do two hypotheses in physics ever constitute such a strict dilemma?”, p. 288; English trans., p.190), and what amounts to essentially the same, when claiming that *reductio ad absurdum* is irrelevant to physics (“unlike the reduction to absurdity employed by geometers, experimental contradiction does not have the power to transform a physical hypothesis into a indisputable truth”, *ibid.*)²¹. Duhem also objects to inductive elimination by arguing that it would work only if the stock of conceivable physical laws were finite, which is impossible. It is easy to see that this is but an iterative generalization of the present argument against crucial experiments.

Actually, this whole line of reasoning seems to be dubious, since it represents the empirical test as a duel between putative laws and observations, which (a) precisely denies to be the correct description. As some have rightly pointed out²², Duhem can use the argument on *reductio ad absurdum* only as a reinforcement of his initial position: *even if, implausibly, tests could be construed as duels*, it would not follow that crucial experiments exist. This may be helpful to add against hard-liner empiricists, but will normally need not mentioning. If the argument has limited dialectical use, this is because it is weaker than the main one through (a). It denies the verificatory component of crucial experiments, which is enough for the conclusion that they do not exist in the strong sense, but simply refuting crucial experiments escape its strictures.

By this logical observation, one may hope to rationalize Popper’s problematic claim that “Duhem in his famous criticism of crucial experiments succeeds in showing that crucial experiments can never establish a theory. He fails to show that they cannot refute it” (1963-1972, p. 112, n. 26). It would seem as if Popper approved of Duhem’s exclusion of *reductio ad absurdum*, with its pleasant antiverificationist consequence, but rejected his argument through (a), which means a challenge to refutationism. If this is indeed Popper’s move here, it is distressing, because he generally accepts the underdetermination thesis, at least in some form. As early as *The Logic of Scientific Discovery* (henceforth *LSD*), he envisaged several reasons why empirical refutations

may not be compelling, in particular mentioning two Duhemian sources, i.e., the presence of auxiliary hypotheses and the equivocality of observations; see the passage on “conventionalist strategies” (1935-1972, §9 and §19-20). To avoid an inconsistency, there is only one interpretation left, i.e., that Popper uses the idea of a refuting crucial experiments loosely. He just means a testing experiment that strongly supports the conclusion that a certain hypothesis – or failing this, a certain compound of hypotheses – is false. But in this diminished sense, Duhem fully accepts that refuting crucial experiments exist, as we now proceed to explain, and Popper has no reason for claiming a serious disagreement with him.

2. 4 Duhem’s answer to the underdetermination problem

We have been careful to distinguish between Duhem’s underdetermination *thesis*, which is a bare statement of the possibilities for refutation, and Duhem’s corresponding *problem*, which is to turn one of these possibilities into a final refutation. As we have reconstructed it, the thesis relies on propositional logic alone and escapes triviality only because of the range of statements that it specifies. With this construal, it does not entail that there is no rational way out of the underdetermination, let alone that there is no logical way out, if logic is taken beyond the propositional realm – these conclusions depend on how the problem is appreciated. Now, Duhem’s answer to the latter is far from being clear-cut, but this is in part of necessity, because it is a pragmatic, and as we will argue, a historical answer.

When certain consequences of the theory are struck by experimental contradiction, we learn that this theory should be modified but we are not told by the experiment what must be changed. It leaves to the physicist the task of finding out the weak spot that impairs the whole system. No absolute principle directs this inquiry, which different physicists may conduct in very different ways without having the right to accuse one another of illogicality. For instance, one [physicist] may be obliged to safeguard certain fundamental hypotheses while he tries to reestablish harmony between the consequences of the theory and the facts by complicating the schematism in which these hypotheses are applied by invoking various causes of error, and by multiplying corrections. The next physicist, disdainful of these complicated artificial procedures, may decide to change some one of the essential assumptions supporting the entire system... Each is *logically* permitted to declare himself content with the work that he has accomplished.

That does not mean that we cannot very properly prefer the work of one of the two to that of the other. Pure logic is not the only rule for our judgments; certain opinions which do not fall under the hammer of the principle of contradiction are in any case perfectly unreasonable. These motives which do not proceed from logic and yet direct our choices... constitute what is appropriately called good sense (“ce qu’on appelle proprement le bon sens”, *TP*, II, VI, §X, p. 329-330; English trans. p. 216-217).

When comparing the conservative and the radical physicists of this passage, an everyday observer might be repelled by the “obstinacy” of the first, or to the contrary, by the “haste” of the second. Duhem does not say which of the two opposite attitudes carries more weight, and does not even suggest that one of them must prevail over the other; there may be a stalemate, as it were. Our interpretation of this passage emphasizes not commonsense, an elusive concept, but the *pattern over time* involved in the psychological description. The conservative physicist becomes unbearable only when he repairs the existing theory *over and over*, and the radical is unpalatable only when he strikes at it *too early*. Here is how Duhem redescribes the conflict in the theory of light:

Biot by a continual bestowal of corrections and accessory hypotheses maintained the emissionist doctrine in optics, while Fresnel opposed this doctrine constantly with new experiments favoring the wave theory.

In any event this state of indecision does not last forever... After Foucault’s experiment had shown that light traveled faster in air than in water, Biot gave up supporting the emission hypothesis... By resisting... for a longer time... Biot would have been *lacking in good sense* (*TP*, II, VI, §X, p. 331; English trans., p. 218).

Putting aside the allusion to commonsense, we find our time-based interpretation reinforced by this quote. Foucault’s experiment came after the same situation had been repeated – one group multiplying reinterpretations, while the other brought out many new facts and hypotheses. The experiment draws its significance from its place in an history. It is *critical* like the turning point of an illness in traditional medicine, a sense which is easy to contrast with the cruciality denied by (b).

More abstractly, we identify Duhem’s answer to the underdetermination problem with the following claims. (i) Sometimes in the history of science, this problem received a satisfactory solution. (ii) These solutions typically arose in a conflictual environment, after a lengthy exchange of arguments in the face of renewed evidence, with the arguments weighing more for one camp from some moment onwards. (iii) These solutions ended up in the qualified rejection of a particular hypothesis, despite the fact that the discussion had involved many other statements and not all of them were checked properly. If Duhem’s final position does include (iii), it goes beyond what is sometimes said of him, i.e., that he took refutation to be capable to hit complex theories and stopped at this conclusion. When Duhem moves from logic to pragmatics and history, we understand him as suggesting that what was eventually refuted in the 19th

century theory of light was the corpuscular theorists' *hypothesis*, i.e., that light is made out of particles, and not simply the *system* in which this hypothesis was encapsulated.

The informality of this resolution stands in sharp contrast with the more definite proposals made by later philosophers of science, especially the Bayesians and the falsificationists. The latter make for an easy comparison because of the refutationist ground they share with Duhem. Popper's own answers to the underdetermination problem follow two rather different lines²³. The first amounts to claiming that the underdetermination problem arises only for those who search for certainty, like the verificationists. Once we are reconciled with the view that all our scientific hypotheses are guesses anyhow, we stop worrying whether or not we blame the wrong subset of premisses for the refutation. This seems very much like throwing the baby with the bathwater. When Popper takes the second line, he must believe that the underdetermination problem matters after all, since he tries to solve it. Essentially, he recommends that the scientist reformulate his premisses so as to make them logically independent of each other, and once he has done so, that he devise a test for each in turn. Unlike the previous one, this is a very substantial suggestion, and indeed, one which scientists have naturally abided by. We will see the rôle it played in the decision theorists' assignment of responsibilities between the different components of VNM theory when the latter became submersed with counterexamples. However, this procedure of *independent reformulation and testing* is clearly not available everywhere: it may fail at either of the two stages, i.e., the logical decomposition of premisses may be lacking in independence, and there is no guarantee anyhow that premisses and tests can be associated in an one-to-one way (we will see that this happened in the VNM case). Moreover, the method is helpful only to spot a falsehood *in the primary hypotheses*, and it works mostly for theories in which these hypotheses have already been axiomatized (VNM theory being one such case). It is unclear how to apply it, except very informally, outside this realm of applications.

About these other sources, Popper has something important to add:

As regards auxiliary hypotheses we propose to lay down the rule that only those are acceptable whose introduction does not diminish the degree of falsifiability or testability of the system in question, but, on the contrary, increases it (1935-1972, § 20, pp. 82-83).

This is excerpted from the already mentioned passage of *LSD* about "conventionalist strategies", i.e., those moves which deflect the refutation from the target hypothesis

when the empirical test is negative. It is in this indirect way that Popper faced the underdetermination problem in his early work. We may generalize his criterion for changes in the auxiliary hypotheses to *all* logically possible modifications of the Duhemian compound. To apportion refutation between the candidate statements, first compare the systems following from each apportionment in terms of their overall refutability. Then, declare an apportionment to be *acceptable* if it results in a system that is more refutable than the initial one and it is not itself refuted empirically. Thus adapted, the rule delivers a crisp necessary condition for an answer to the underdetermination problem. Comparison with Popper's later work, especially *Conjectures and Refutations*, suggests that it is well representative of his long-run thinking on this problem. It agrees with his general claim that *a necessary condition for scientific progress is that the new theory will have more testable content than the current one*²⁴. We may add that a *best* apportionment within the acceptability class is one which results into a maximally refutable system, in the sense that no other acceptable apportionment generates a more refutable system. Again, this agrees with what Popper argues about scientific progress.

The straightforward objection to these suggestions is that they are too strict. To illustrate by an example loosely related to EUT, it would not permit replacing the false regularity:

(1) "All individuals, in all their dealings, are risk averse".

by, say, the following one, which is not so obviously false:

(2) "All individuals with a low income, in all their dealings except those concerned with gambling with small stakes, are risk averse".

Despite the sharp restriction of scope, hence the decrease in refutability, it is hard to justify that generality (2) should be excluded from consideration. On further examination, it may turn out to involve an optimal mix of content and approximate truth among the variants of (1). To define an acceptable apportionment in the way just said will block (2) from competing for the status of a best revision. We elaborate here on an objection made by Lakatos (1970, p. 182) to the effect that Popper disallows restrictions in the antecedent of well-established laws that serve to absorb anomalies. This is too specific a way of putting the objection, because the need for introducing restrictions of false generalities can arise from many sources. If it is absurd to reject them off hand, this is simply because they may be the only true generalities that are to be found.

Lakatos is still too much of a Popperian in order to accept the possibility that empirical sciences can make genuine advances *while losing in overall refutability*, and he eventually bans it from the definition of progress that regulates his “methodology of scientific research programmes”²⁵.

Although himself a refutationist, Duhem gives no sign of having ever envisaged Popper’s and Lakatos’s ban. If our EUT case study carries any normative force, it will suggest that Duhem was right, and the two falsificationists wrong. For decision theorists groped their way from the refuted VNM theory towards alternatives that involved losses of content. This is one of the several challenges posed by their work, which we now proceed to explain.

3. THE FORMATIVE STAGE OF VON NEUMANN-MORGENSTERN THEORY

3.1 *Von Neumann and Morgenstern’s contribution to expected utility theory*

This case study is concerned not with expected utility (EU) theory as a whole, but only with its more elementary part, which is referred to as *von Neumann-Morgenstern* (VNM) *theory*. The standard exposition consists of a small body of axioms – usually three and at most five – followed by a few salient consequences, the most important one being the VNM *representation theorem*. According to the latter, if an individual’s preference relation is defined on risky prospects involving given probabilities (“lotteries”) and conforms to the axioms, this preference relation gives rise to a numerical function having the *expected utility form*, i.e., the form of a sum of the utility values of outcomes weighted by the probabilities of these outcomes. The function *represents* the individual’s preference relation in the sense that, for any pair of lotteries, it gives a higher value to the one which is strictly preferred to the other (and it gives equal values to indifferent lotteries). This is the *existence* part of the VNM representation theorem. The *uniqueness* part says that the EU form of the representing function identifies it uniquely up to positive linear transformations. Exact statements will be given in the next subsection.

VNM theory is a paradigmatic example for contemporary choice theories. All these theories rely on a sharp distinction between information relative to (an individual’s)

preference and information relative to (this individual's) *utility*. Only the former is the substantial concept; the latter is but a numerical index representing, in the technical sense explained, the agent's preferences or choices (for current purposes, we may follow the questionable habit taken in the field of disregarding the significant difference between these two notions). This conceptual hierarchy sets a target for an axiomatization of the choice theory. The symbols appearing in the axioms should concern either the objects of preference or the preference relation, while the symbols of utility should be reserved for the logical consequences of the system, such as the representation theorem. This very clear structure became pervasive in the 1950's under the influence of von Neumann and Morgenstern (1944-1947), the eponymic founders of the theory we are concerned with²⁶.

VNM theory takes as given the probabilities from which it derives the EU representation, so its limited task is to derive a utility function on the outcomes and combine this output in the EU way with the preexisting probabilities. A primitive concept in the VNM axiom system, the *lottery* is a mapping associating probability p_1 with outcome x_1 , probability p_2 with outcome x_2 , and so on; abstractly, it is a probability measure p on the set of outcomes X . A more sophisticated theory, Savage's (1954-1972), *derives* the probability measure p at the same time as the u function, and shows it to be unique given the axiomatic system. The latter replaces lotteries by other objects of preference, the *acts*, which involve no prior quantitative information. Many, especially in economics, contrast the two branches of expected utility theory as *objective* and *subjective*, respectively, but from the philosopher's point of view, these are misleading expressions. Savage's endogeneous probability measure can be interpreted as being subjective, in the sense that they express the agent's degrees of belief, but there is nothing in von Neumann and Morgenstern to warrant an objective interpretation of the lotteries, even in a less specialized sense of "objective" than frequentism. There is no more to be said in general of these objects of preference than that they are *given*. Admittedly, they often refer to actual lottery tickets with a physical device underlying them, such as those sold by the National Lottery, but they may also represent combinatorial possibilities in a game of chance, or an agent's subjective probabilities, however this agent arrived at these estimates. In brief, VNM theory is too terse to allow for a single interpretation, a warning that also applies to the hypotheses eventually proposed to replace it.

In *Theory of Games and Economic Behaviour* (1944-1947), von Neumann and Morgenstern gave the impetus to the theory labelled after them, but their technical treatment differs from that which came to be received afterwards. Paradoxically, the book puts in place the style of axiomatization now received in choice theory without fully abiding by it. Chapter I,3 of the 1944 edition and the 1947 Appendix developing this chapter are expressly devoted to “the axiomatic treatment of utility”. Contrary to what this expression would now suggest, von Neumann and Morgenstern do not take the preference relation on lotteries as a primitive concept. Instead, they start from comparisons bearing on equivalence classes of lotteries, as are defined by the indifference relation. Their work consists in axiomatizing comparisons made between these peculiar objects, and from these comparisons, demonstrating the existence and relative uniqueness of the EU representation.

The previous discussion may seem to be abstruse, but it goes a long way to account for the strange start made by VNM theory. The truly substantial axiom of this theory - that which was to matter dramatically to both its empirical testing and normative assessment - is the so-called *von Neumann-Morgenstern independence condition*. But remarkably, it came to be identified only after the founders, and this is because they had hidden it behind their technical choice of objects of comparisons. This slip was pointed out by Samuelson and Malinvaud rather belatedly - at a conference held in Paris in 1952, which was to have other lasting consequences that we will soon review²⁷.

At that time, a reasonably explicit set of axioms, due to Marschak (1950), was already available. Then came the entirely transparent ones of Friedman and Savage (1952), Samuelson (1952 a and b), Herstein and Milnor (1953), Luce and Raiffa (1957), and still others. Each of these works obeys the pattern of today’s choice theory by moving from the preference relation on lotteries to the EU representation, and each includes a relevant version of the - ironically called - VNM independence condition.

3.2 *An axiom system for VNM theory*

We will discuss the axiomatic work in terms of what may be the easiest system of all, i.e., Friedman and Savage’s (1952). Its axioms strongly overlap in logical content, which is a defect from the logician’s perspective, but makes it more tractable than more independent systems like Herstein and Milnor’s (1953). A *lottery* is defined to be a

probability measure p on a set X , called the set of *outcomes*. Assuming finiteness for convenience, we can write

$$X = \{x_1, \dots, x_i, \dots, x_n\} \text{ and } p = (p_1, \dots, p_i, \dots, p_n) \text{ with } p_i \geq 0, \sum p_i = 1$$

(that is to say, probability measures are simply probability vectors). A *weak preference* is a binary relation R on the set L of all such vectors, with “ $p R q$ ” reading as “ p is strictly preferred or is indifferent to q ”. The system consists of three axioms on R and L :

(A1) (Ordering) R is an ordering (i.e., it is reflexive, transitive, complete) over L .

(A2) (Continuity) R is continuous²⁸.

(A3) (VNM Independence) For all p, q, r in L , and all numbers a between 0 and 1 (0 excluded), $ap + (1-a)r R ap + (1-a)r$ if and only if $p R q$.

The VNM representation theorem follows. Formally: if (P1), (P2), and (P3) hold, then

(Existence) there exists a numerical function u on X such that

$$(*) p R q \text{ if and only if } \sum p_i u(x_i) \geq \sum q_i u(x_i),$$

and

(Uniqueness) the functions u' on X satisfying the same property (*) as u are exactly the functions $u' = \alpha u + \beta$, where α is a positive number and β any number.

Both axioms (A2) and (A3) involve convex combinations of elements in L that one would like to interpret as *compound lotteries*, i.e., lotteries the outcomes of which are lotteries, not final outcomes. However, mathematically, a convex combination of probability vectors is another probability vector, which means that the chosen formalism cannot express the distinction between compound and simple (non-compound) lotteries. Take for example:

$\underline{p} = (5 \text{ M FF with prob } 0.10, 1 \text{ M FF with prob } 0.89, 0 \text{ FF with prob } 0.01)$, and

$\underline{p}' = (1 \text{ M FF with prob } 0.89, r \text{ with prob } 0.11)$, where $r = (5 \text{ M FF with prob } 0.10, 0 \text{ FF with prob } 0.01)$.

Once the relevant multiplications are made, \underline{p} and \underline{p}' become the same 3-dimensional probability vectors. So the formalism leaves no choice but to identify the corresponding lotteries, which is objectionable because the agent may react differently to them. Luce and Raiffa (1957, p. 26) address this problem by redefining lotteries as lists of *prizes* and corresponding probability numbers, each prize being a lottery of lesser complexity, which may be an outcome as a particular case. This formalism allows for much flexibility in the description of the agent's preferential attitudes. Luce and

Raiffa do not pursue the topic, and are content with imposing the axiom of *reduction of compound lotteries*, to the effect that the agent reacts in the same way, whatever the stage of reduction of the lotteries presented to him. This leads back to the standard assumption, but makes it explicit, instead of burying it in the definition of a mathematical primitive. The higher degree of explicitness facilitates the interpretation and testing of the overall system, so we will take Luce and Raiffa's axiom to be a component part of VNM theory, even if we do not state it formally²⁹.

Following this empirical line, we list a few situations in which the supplementary axiom would be violated. (a) The compound lottery stands for a physical device that does not satisfy the probabilistic independence of the successive stages, so that the associated simple lottery cannot be computed without more probabilistic information being given. (b) In the probabilistic independence case, the individual does not know how to make the calculations. (c) The individual feels a specific interest in participating in a many-stage game rather than in playing just once. To adapt an example from Ellsberg (1954, p. 543-544), a player in a sequential poker game might choose to stay in this game even though he would not accept the one-stage game with the calculated multiplicative probabilities. This gambler likes being involved, and this is why he stays in despite a hand which he would otherwise judge to be unpromising. The example falls under the heading of *the pleasure of the game objection*, which VNM theorists have acknowledged unproblematically. By contrast, they pay little or no attention to objections (a) and (b), perhaps because many concrete examples of lotteries do satisfy the probabilistic independence assumption denied in (a), and because (b) is just another intractable example of bounded cognitive ability. Whatever the differences of theoretical attitude vis-à-vis (a), (b) and (c), many would regard failures of the reduction of compound lotteries as constructive suggestions on how to restrict the domain of application of VNM theory appropriately.

3.3 Ordering and continuity

By a well-known result in choice theory, if a preference ordering on some set of alternatives is continuous in a suitable technical sense, there exists a continuous numerical function on this set representing the preference, and it is unique up to monotonically increasing transformations. Comparing this statement with (A1) and

(A2), on the one hand, and the VNM representation theorem, on the other, one can see what the contribution of (A3) consists of. Essentially, (A1) and (A2) deliver a representation V of R on L that is left unspecified, except for the weak properties of being continuous and equivalent only to its monotonic transforms (like V^2 or $\text{Log } V$, and unlike $-V$ or $1/V$). Adding (A3) has the effect of restricting both the functional shape of V (by conforming it to the EU formula, or in mathematical terms, making it linear in the probabilities) and the class of its admissible transformations (V is then equivalent only to monotonic linear transforms). That it took time to isolate (A3) can partly be explained by the fact that early VNM theorists were not yet accustomed to the mechanics of representation theorems³⁰.

How are the two axioms to be assessed? Following a widespread view in both choice theory and microeconomics, (A1) states a minimal requirement of rationality, thus being unobjectionable normatively, and it is also a roughly plausible empirical generality. Expected utility theorists have most of the time refrained from questioning it, and we will not discuss it in this paper. Concerning (A2), the predominant view is that it is a technical condition without any clear empirical counterpart serving as either evidence or counterevidence. Interestingly, however, not every VNM theorist has followed this non-committal line. Marschak (1950, p. 117) demonstrates that in the presence of the ordering property and a relevant part of VNM independence, his version of continuity leads to the following:

(C) For all p, q and a strictly between 0 and 1, if $p P q$, then $p P ap + (1-a)q P q$.
(P stands for the *strict* preference relation that is induced by R .)

This is tantamount to saying that an individual faced with a compound lottery having two outcomes will not value this lottery more than the outcome he values more, and not value it less than the outcome he values less. Thus, (C) sounds like an acceptable conclusion, both normatively and empirically. But Marschak puts forward an intriguing counterexample that is worth reconsidering in detail.

A typical mountain-climber dislikes being alive for sure ($= p$) and incurring death for sure ($= q$), ranking the former of these two (here degenerate) lotteries above the latter, whereas he likes risking his life with a 95% chance of surviving the climb ($= ap + (1-a)q$). That is to say, he strictly prefers a compound lottery to *both* of its components, and not only to the less preferred one, thus violating (C). Marschak's (1950, p. 138-140) comment stresses that the same "love of danger" can also underlie activities – like the

funding of scientific research or the undertaking of geographic expeditions – that are standardly classified as being “economic”. The proper distinction, Marschak goes on, is between *two different ways of taking chances*, and it will typically cut across the same spheres of human activities. So far, so good; but then, surprisingly, Marschak does not try to identify the two ways in terms of relevant psychological, behavioural, or situational properties. Instead, he discusses the sense in which VNM theory can account for chance-taking action.

Given a suitable set of outcomes X (the easiest example being a set of monetary amounts), the theory defines an individual to be a *risk-lover* if the u function delivered by the VNM representation theorem is convex, and a *risk-avorter* if it is concave. The concept of risk-love is the single tool made available by VNM theory to explain gambling, lack of insurance, speculation, and the like, and it works well on many activities, but it also fails on some, and the mountain-climber is a sign-post for these exceptions. The turn in Marschak’s discussion suggests that it must have been difficult, even perhaps impossible, for him to assign a domain to VNM theory *from the outside*. He resorts to the theory *itself* to say where this theory applies. His argument is not vacuous because he makes two relevant points: firstly, that exceptions of the rock-climber sort are essentially unmanageable for the theorist (a violation of (C) excludes the existence of a utility representation, whether VNM or not), and secondly, that they are best regarded as being irrational.

We have gone through this intriguing discussion because it reveals salient features of VNM theory – and perhaps even of choice theories – at large. First, it confirms that the theory was accompanied with recognized counterexamples from its very beginning, these counterexamples being regarded as not particularly threatening. The tendency was to view them as suggesting domain restrictions rather than stating refutations or even anomalies. For historical reasons that we cannot develop here, *the pleasure of gambling* functioned like a convenient buffer zone for various problematic cases³¹. We have seen this concept at work in Ellsberg’s discussion of Luce and Raiffa’s compound lottery axiom. Marschak inherits its defensive use and attempts to refine it by separating those cases of the gambler’s pleasure which VNM theory can redeem and those which it cannot, and here we record a second noteworthy point. Instead of invoking the nature of the chance-taking actions, the theorist directs attention to the *theoretical tractability* and the *rationality*, as assessed by himself, of these actions. From now on, the evaluation of

empirical performance of VNM theory will be mixed with an assessment of these qualities, despite what may seem to be an irrelevant association. Third, the theorist does not have a preexisting yardstick of rationality, but argues about it in a circle of reinforcing considerations. Daring mountain-climbers suggest an initial diagnosis of irrationality that is being confirmed by the technical observation that their preferences clash with the axioms, and in the particular instance, with any smooth mathematical description. The tractability argument here comes to the rescue of the rationality attribution.

3.4 *Early theoretical attitudes towards VNM independence*

With the benefit of hindsight, one would have expected the very substantial independence condition to become the focus of theoretical attention once it was recognized, but this was very far from being the case. Those who filled the gap in von Neumann and Morgenstern's axiomatic system believed that it was an unproblematic addition to make. The first to formalize independence, Marschak wrote that it was "very weak" (1950, p. 111) and spent less time on it than on continuity. At an already late stage, Ellsberg (1954) offers a stunning example of this general neglect. Having recalled the entrance of (A3) into the VNM axiom system, he claimed that "it seems rather hard to justify this emphasis since the axiom seems indubitably the most plausible of the lot" (1954, p. 544), and, like Marschak, focussed his discussion, both normative and empirical, on the other conditions³².

At least Friedman and Savage (1952, p. 468-469) and Samuelson (1952a, p. 133-134) felt the need for a justification of (A3). The former claim that (A3) follows from an elementary principle of choice theory, the *dominance principle*, which in the particular context would go like this:

If an agent weakly prefers lottery p to lottery q in case event E occurs, and weakly prefers lottery r to lottery s in case the complementary event E^C occurs, then this agent should weakly prefer the prospect giving p in case of E and r in case of E^C to the prospect giving q in case of E and s in case of E^C .

Virtually no decision theorist objects to the reasonableness of this principle, so it would buttress (A3) if this condition did logically follow from it, but this is not the case. The dominance principle relates to *prospects*, which are defined vis-à-vis states or events,

not lotteries, which are stated in terms of probabilities, and this turns out to make a world of difference. The dominance principle can be reformulated to fit VNM theory, but when this is done, it turns out that (A3) implies it, and not the converse³³.

As for Samuelson, who boasted having become a convert to VNM independence once he knew how to formulate it, his justification amounts to the following. Consider two compound lotteries $l = ap + (1-a)r$ and $l' = aq + (1-a)r$, and let us ask how the agent would assess them at the *interim* stage, that is, after these lotteries have been drawn, resulting in either p for the first and q for the second, or r for both of them, and before these lower-level lotteries are drawn. At this stage, the choice between l and l' becomes one between p and q , since there is nothing left to choose in the case where r has been drawn. Taking for granted that the ex ante preference between l and l' should conform to what the interim stage will be, Samuelson concludes that (A3) is warranted. Contrary to Friedman and Savage's, this argument initiates a conceptually fruitful line. It foreshadows the later *dynamic analysis of EUT*, in which many today see the strongest possible normative arguments that can be made for VNM independence (or the related condition in Savage)³⁴. However, as this more recent work makes clear, more than one assumption must be made for Samuelson's argument to be logically complete, each being normatively questionable. The brief and authoritative comment of the 1950s does not do justice to the complexity of the logical and conceptual issues it raises.

3.5 A holistic argument for VNM theory

There was another course to take for those early writers, mostly economists, who were trying to protect the budding VNM theory against premature dismissal. Instead of discussing individual conditions or predictions in the face of partial evidence, they could, despite the still limited empirical record, inquire about the performance of the theory as a whole. Equally, instead of trying to assess it in isolation, they could place it against the background of earlier treatments of choices under risk in economics. This double shift towards holism, with essentially soothing conclusions, can be found in the first paper by Friedman and Savage (1948), which forms an interesting contrast with the second.

The two writers go back to Marshall's *Principles* (1890-1921), a treatise which was still influential in Anglo-American quarters at the time. The passages of this book which discuss gambling dash any hope of handling any behaviour of this kind, even of an economically relevant one, with the tools of economic theory³⁵. Marshall's argument is already couched in terms of expected utility calculations, which he performs under the assumption that the u function is concave – a reflection of his intangible principle that marginal utility is decreasing. With this assumption, any form of gambling involves a loss, insurance always involves a gain, and it is therefore irrational to gamble and not to insure. The final implication is that if economics is to remain the science of rational conduct, there is no place in it for an analysis of gambling and – more importantly, of course – of the broadly similar economic activities in industry, commerce and finance. This appears to be a severe self-limitation of the discipline.

Friedman and Savage (1948, p. 280-281) contrast von Neumann and Morgenstern's work with this discouraging Marshallian line, which they correctly take to be typical of earlier economics. They emphasize the technical distinction that we introduced when discussing Marschak, between the case of *risk-love*, corresponding to that of a convex u function, and the case of *risk-aversion*, corresponding to that of a concave u . These two possibilities are equally compatible with the individual's preferences satisfying the VNM axioms, which, and which alone, constitute the VNM rationality criterion. The step accomplished by *Theory of Games* is to redefine "the problem of rational behavior" in such abstract terms that the concavity of u is no longer an analytic component of its solution, as it was in Marshall. A major consequence for the progress of economics, it now becomes possible to maintain that it is the science of rational conduct and nonetheless subject several forms of gambling and related activities to its theoretical investigations.

Perhaps mostly with a view of illustrating the new possibilities, Friedman and Savage propose, as a typical shape of utility function on monetary amounts, that u is concave on both small and large amounts of money, and convex in between. They suggest that this hypothesis "rationalizes" (1948, p. 293) – presumably, in the two senses of explaining and making rational sense of – the broad fact that the same individuals gamble and take out insurances, although for different amounts of money and different levels of income. On the face of it, this was an empirical hypothesis to be tested, and there was a detailed experimental study by Mosteller and Noguee (1951) –

one of the first ever made on EUT – which concerned itself with the proposed shape of u . The broad conclusion was that the suggested shape was not supported by the data³⁶. Despite its hand-waving style, the hypothesis deserves to be remembered as an example of the new theoretical standpoint that emerged in the early 1950s.

We may conclude as follows on the formative stage of VNM theory. Prior to it, economics had little, if anything, to say about choices under risk in general. These choices could not be investigated properly because, by their very nature, they encompassed the excluded case of gambling-like behaviour as a relevant possibility. All of a sudden, thanks to von Neumann and Morgenstern, economists felt capable of exploring insurance behaviour, portfolio choices, job search, the firms' or the government's policies in a risky environment, and similarly important topics. Perhaps even more than the first readers of *Theory of Games* could have guessed, the twenty or thirty years following the book witnessed a continuous flow of deductions, usually from the simple VNM groundwork, but sometimes also from more refined theories like Savage's. Thus, with appropriate initial conditions and auxiliary assumptions, VNM theory was shown to entail definite statements on the agents' degree of insurance coverage, leading at long last to some elements of insurance theory³⁷. A few – not all – of these entailments can be construed as empirical predictions, which testifies to the leap forward made in both logical and testable content. In retrospect, it seems difficult not to approve of the strategy pursued by the contemporaneous economists, like Marschak or Friedman and Savage, of protecting the fragile developments from *prima facie* counterexamples and foundational queries.

4. THE CRITICAL STAGE OF EXPECTED UTILITY THEORY

4.1 *The 1952 turning point*

The critical stage of EUT can be dated back to some events of the year 1952, after which the field was forever divided between its supporters and adversaries. First of all, a major conference gathering both North-American and European specialists of decision theory took place in May in Paris³⁸, and intense controversies burst out within this distinguished attendance, which divided into roughly three groups. The American

school, led by Marschak, Samuelson and Savage, was pushing the case for VNM theory all the more energetically since it could now rely on the sophisticated EU variant that the Savage had devised for subjective probabilities³⁹. In this campaign, they received the idiosyncratic support of a founder of the subjective probability school, de Finetti. They were a powerful team, and in retrospect, one is struck by the boldness of the French economic engineers – most prominently, Allais – who pugnaciously objected to EUT throughout the conference. The majority of remaining participants were econometricians or mathematicians, who acted like witnesses – occasionally, like arbitrators – in the conflict between the bull-doing majority and the restless minority.

The year 1952 witnessed a second event of significance, which is not as straightforward to locate as the first. From then on, VNM theory would be discussed primarily in terms of *a privileged format*: a subject is asked to choose in succession between two pairs of lotteries with monetary outcomes, i.e., (x_1, y_1) and (x_2, y_2) ; the probability and outcome values are fixed in such a way that if the subject obeyed the VNM theory, he would choose x_2 if and only if he chose x_1 . The two-pair format is especially relevant to the testing and normative appraisal of the VNM independence condition (A3), whereas we have seen that a single-pair format is sufficient to discuss the continuity condition (A2)⁴⁰. Strange though it seems, the first sustained empirical attempts at testing VNM theory did not use the apparently obvious check that consists in presenting subjects with two (or more) relevant pairs. Rather, they implemented the following sequence: first, let subjects choose between lotteries $((p, x), (1-p, 0))$ with variable money outcomes x and probability values p ; second, estimate utility functions $u(x)$ from the choice data; third, test VNM theory by checking whether $Eu(x)$ accounts for these or (preferably) other data sufficient well. Why this “painstaking procedure” – the expression of a contemporary experimenter, Camerer (1995, p. 621) – lingered on to such an extent is a question we leave for the final assessment section.

The most famous (and indeed a very telling) example of a two-pair experiment is the so-called *Allais paradox*, which occupied centre stage in the experimental and theoretical developments of the 1970s and 1980s. The word “paradox” suggests connections with upsetting discoveries such as Russell’s paradox in set theory or the EPR paradox in physics, but there is no deep significance to be attached to it now; it is used simply because the subjects in Allais’s experiment usually choose against VNM theory, the established doctrine - the “doxa” - of the field. The history of this example is

a genuine curiosum⁴¹. It occurred for the first time during the 1952 conference, but is reported nowhere in the proceedings⁴². According to a story which is now part of the decision theorists' folklore, Allais asked Savage to choose from two pairs of lotteries when they were lunching privately, and his astutely chosen figures trapped him into a violation of VNM theory. Savage contributed to popularize the episode by revisiting it in his *Foundations of Statistics* (1954-1972); we will see how he made things worse for himself by trying to explain his blunder. The same figures appeared in Allais's 1953 *mémoire* and the abridged journal form of this piece, also out in 1953, which is now the standard reference for the paradox.

Allais tried to debase VNM theory in two further ways, one of which is another paradoxical application of the two-pair format, and the other follows a different pattern, in which the damaging consequence is an inconsistency in measurement. These findings also belong to his work of the years 1952 and 1953. All in all, decision theorists were more impressed by the critical insights of this work than by the effort it also made to build a positive theory of risky choice. Following this line, we will discuss Allais's paradox and the two lesser known paradoxes at some length. We then proceed to explain how they were turned in the 1970s and 1980s into the genuine experiments which they initially were not, and the rest of the section reviews some of the non-VNM hypotheses of the time, with a view to prepare the final assessment of VNM theory.

4.2 *The Allais paradox*

In the journal article resulting from his *mémoire*, Allais (1953b, p. 526-527) claims to have elicited answers to the following two questions:

Question 1: Which lottery would you choose of $x_1 =$ to receive 100 millions FF with probability 1, and $y_1 =$ to receive 500 millions FF with probability 0.10, nothing with probability 0.01, and 100 millions FF with probability 0.89?

Question 2: Which lottery would you choose of $x_2 =$ to receive 100 millions FF with probability 0.11, and nothing with probability 0.89, and $y_2 =$ to receive 500 millions FF with probability 0.10, and nothing with probability 0.90?

The numbers are so devised that a subject obeying VNM theory should choose either x_1 and x_2 , or y_1 and y_2 . For there is no numerical function $u(x)$ that would both satisfy the inequalities corresponding to the choice of either x_1 and y_2 , or of y_1 and x_2 :

$$u(100) > (\text{resp. } <) 10/100 u(500) + 89/100 u(100) + 1/100 u(0)$$

and

$$11/100 u(100) + 89/100 u(0) < (\text{resp. } >) 10/100 u(500) + 90/100 u(0).$$

Most very cautious people (“la plupart des gens très prudents”, *ibid.*), Allais contends, will choose x_1 and y_2 .

As Allais thought of the example, it affected (A3) rather any other part of VNM theory. To clarify the further step involved here, let us introduce the auxiliary lottery z : $z =$ to receive 5 millions with probability 10/11, and nothing with probability 1/11, and then reformulate the experiment as follows:

Question 1': which lottery would you choose of $x'_1 =$ to receive 100 millions FF with probability 11/100 and 100 million FF with probability 89/100, and $y'_1 =$ to receive lottery z with probability 11/100, and 100 millions FF with probability 89/100?

Question 2': which lottery would you choose of $x_2 =$ to receive 100 millions FF with probability 11/100 and nothing with probability 89/100, and $y'_2 =$ to receive lottery z with probability 11/100 and nothing with probability 89/100?

The newly introduced lottery x'_1 and y'_1 have a common outcome occurring with a common probability – a *common consequence* in the later terminology –, and so do x_2 (this lottery is unchanged) and y'_2 , the common probability being identical in the two pairs. Now, (A3) says that it is inessential to an individual's preference how a common consequence is fixed; hence, he should strictly prefer x'_1 to y'_1 if and only if he strictly prefers x_2 to y'_2 , and what he violates in VNM theory is precisely (A3). To draw this conclusion for the initial design requires one's accepting *the axiom of reduction of compound lotteries*, which turns y'_1 and y'_2 into equivalent forms of y_1 and y_2 (the rewriting of x_1 into x'_1 is unproblematic). Allais did not have to worry about the other axioms (A1) and (A2), which cannot be violated in this two-pair example.

Today's retrospectives often distort the initial intent of the paradox by describing it as a straightforward empirical test, which it was not and could not be from Allais's perspective⁴³. When he first wrote about it, he was unable to document the prevalence

of each answer; presumably, he had just experimented casually, as in the Savage episode. Besides, his article emphasizes, and it will then become a theme of his work, that *only rational peoples' answers matter*. Whatever evidence Allais might have gathered, the article claims it only for subjects who are, for one hand, rational by common consent (1953b, p. 525, 528), and for another, well acquainted with the probability calculus (p. 524). It is because of the last assumption that Allais takes the compound lottery axiom for granted, apparently not noticing that this axiom can fail for other than cognitive reasons. He did undertake more thorough empirical research than his 1953 article suggests, but always limited to subjects exhibiting the above features, in practice a handful of high-brow students and colleagues⁴⁴.

Putting the paradox further in context, we find that Allais employs it *as a step in a normative argument*. Throughout the Paris conference, he strongly assumed that the EU hypothesis was false as an empirical generality, and even more strongly, claimed that his theoretical opponents shared this view, thus narrowing the disagreement with them to the issue of rationality.

Everybody recognizes the fact that man in reality does not behave according to the principle of Bernoulli [= the EU hypothesis]. There does exist a profound difference, however, in points of view as to how a rational man ought to behave. According to the American school, a rational man must conform to the principle of Bernoulli. In our view, this is a mistake (1953b, English summary, p. 504).

In order to advance the remaining issue of rationality, Allais – and this will also recur in his work – developed two strategies simultaneously. One amounts to arguing against EUT, and especially VNM independence, from the very meaning of the rationality concept. The other consists in *observing what rational men do*, and it is at this juncture that he introduced the paradox. Given the special argumentative purpose, the lack of empirical detail to fill it out is perhaps not surprising. The experiment, if this concept applies at all here, is *a thought experiment*, and it is arguably not even an empirical test, just the hint of it. The point is worth emphasizing, not only because it refines the received history of EUT, but because it creates an interesting tension for the philosophy-of-science reconstruction to come. We will have to check whether a Duhemian account can possibly register Allais's focus on the normative.

One should not make too much of the difference between the semantic argument about rationality and the thought experiment properly. Since the latter leaves undefined

what characterizes “a man considered as rational”, one naturally turns to the former for clarification. Allais offers an “abstract definition of rationality” hinging on two main components: ordering and the dominance principle that was introduced in last section (roughly speaking, if, compared with l' , l increases the probabilistic weight of the better outcomes at the expense of the worse outcomes, l is preferred to l')⁴⁵. Arguably, this minimalist definition fits in both with commonsense and the tradition of economics, which insists on not stretching rationality beyond mere consistency (1953b, p. 519). With rationality so conceived of, the VNM independence condition appears to be a gratuitous addition.

There is another line about *prudence*, which should be kept distinct from the argument about rationality, because it is not so much normative as psychological. Evidently, someone who selects x_1 from the first pair must be very prudent since he gives up x_2 , which has a much higher expected monetary value, just because of a 1/100 chance in this lottery of getting nothing. VNM theory would rationalize this choice by attributing an extremely concave function $u(x)$ to the subject; but then, $Eu(x)$ would predict the selection of y_1 in the second pair. That is, for this theory, the difference that a 1/100 chance of 1 million FF makes to the subject is the same whatever the other chances and outcomes may be. By maintaining that a cautious subject can select y_2 without falling into an inconsistency, Allais both rejects the standard theorizing in terms of expectation and concavity, and makes room for a novel conception in which *prudence would manifest itself variably depending on the particular pair of lotteries*. The same 1/100 chance of 1 million FF does not weigh the same when certainty is an option (first pair) and when both options are uncertain (second pair). In other, still Allaisian terms, there are *psychological complementarities between the chances*, which may be very strong when chances add up to certainty and quite weak otherwise.

The supporters of VNM theory were unshaken by these arguments. Unlike Allais, they did not have, nor wished to have, an explicit definition of rationality, and essentially reproduced Marschak's – after all classic – mode of reasoning in semantic circles from rationality conditions to examples, and vice-versa. Presumably, they thought that rationality was too elusive a concept to be clarified in the heavy style of a French engineer. It could only be discussed at the level of elaborate intuitions, and since these were far from being decisive, extraneous considerations, such as tractability and economy of description, had to be resorted to. Thus, it was seen as a decisive advantage

of VNM theory that it summed up the individual's risk attitude in the shape of his $u(x)$ function, while preserving the convenient linearity of V in terms of probability values. This argument had widely circulated even before it served as defence against the Allais and related counterexamples⁴⁶. In the *Foundations* (1954-1972, p. 101-103), Savage tried to buttress the orthodoxy from still another side, claiming that only reflective choices counted normatively and that the paradoxical answers were no such choices⁴⁷. Himself trapped by Allais's questions, he claimed to have restored preferences satisfying (A3) when he had noticed that he was violating it. He had made a mistake and corrected it; his thoughtful revision was an indication that he was at last answering rationally. But if a thoughtful revision points towards rationality, so does a thoughtful adherence to the paradoxical choice under the same informational circumstances, and Savage's alleged condition for rationality may well work like a boomerang.

4.3 *Experiments on the common consequence and common ratio effects*

The Allais paradox would have remained a theoretical curiosum if it had not been for the efforts of a few decision theorists and psychologists to turn it into something like a controlled experiment. The first studies to be carried out, in the late 1960's and the early 1970's, involved asking variants of Questions 1 and 2 to a small group of subjects – typically, university students, but sometimes business people. The upshot was that the paradoxical choices occurred in 27% to 42% of the cases (these figures apply nearly as well to the violation of VNM theory, since the choice of x_2 and y_1 hardly occurred). One study allowed a subgroup of subjects to make their choices while discussing them, another involved the subjects' reconsidering their choices in view of the pros and cons stated by the experimenter, and a third had each subject assess the choices of others. In this way, the experimenters were paying some attention to Allais's concern about rationality. Above all, they were turning Savage's criterion of normative adherence against his favourite theory, for the paradoxical answer was the most frequent response in all types of sessions and even turned out to be slightly reinforced in one experiment involving reconsideration⁴⁸.

A sophisticated study was later realized by MacCrimmon and Larsson (1979) to double-check both the empirical prevalence of the paradox and its robustness to critical reflection. On the former score, they replaced some of the numerical figures in

Questions 1 and 2 by *parameters*, which they let vary on a range of values that preserved the abstract structure. (I.e., Question 1 is the same as Question 2 but for a change of common consequence: see the analysis of last section.) This resulted in a rate of violation of about 30%, thus close to the initial findings. On the latter score, MacCrimmon and Larsson introduced the concept of a *rule* in order to capture the subjects' underlying motivations. Stated abstractly, but in natural language, rules would offer an educated subject a means of ratiocinating about his choices. Some were mediating the formal conditions of VNM theory, while others generalized the paradoxical answers directly. The idea was to compare the subjects' answers to choice problems with *their approval or disapproval of rules bearing positively or negatively on these choices*. Thus, MacCrimmon and Larsson asked their 19 subjects to give a score to each rule they presented, and eventually classified them into a consistent and an inconsistent group. Roughly speaking, with a consistent subject, the rules he obeys score better than the rules he violates, and the opposite with an inconsistent subject. The first group deserves more attention than the second, since its choices appear to have been more thoroughly reflected (and on Savage's criterion, to be more rational). In this privileged group, MacCrimmon and Larsson (1979, p. 368) noted a significant minority of paradoxical choosers.

This result may seem to be only mildly unfavourable to VNM theory; after all, a majority of consistent subjects were implicit EU maximizers. But if one interprets Allais's claim against the theory as being a barely *existential* one, as the context of his discussion makes plausible, this half-baked result is more than is needed. Besides, it should be recalled that Allais expected paradoxical answers only from *the very cautious* among the rational people. In other words, the majority of MacCrimmon and Larsson's consistent group can possibly comprise of individuals who would depart from VNM theory in circumstances where a paradoxical choice would not require such a large amount of prudence as in the Allais-style examples.⁴⁹

Allais's *mémoire* and ensuing paper (1953b, p. 529) contains another thought experiment that is no less instructive than the celebrated paradox.

Question 3: Which lottery would you choose of $x_3 =$ to receive 100 millions FF with probability 1,
and $y_3 =$ to receive 500 millions FF with probability 0.98 and nothing with probability 0.02?

Question 4: Which lottery would you choose of $x_4 =$ to receive 100 millions FF with probability 0.01 and nothing with probability 0.99,

and $y_4 =$ to receive 500 millions FF with probability 0.0098 and nothing with probability 0.9902?

Arguably, some rational and prudent people choose x_3 and y_4 . This is another violation of VNM theory and – Allais contends again – of (A3) specifically. In view of the compound lottery axiom, the example is restated as:

Question 3: unchanged,

Question 4': Which lottery would you choose of $x'_4 = x_3$ with probability 0.01 and nothing with probability 0.99,

and $y'_4 = y_3$ with probability 0.01 and nothing with probability 0.99?

As it now appears, the violation of (A3) is obtained by offering in the second pair each alternative of the first with the same chance 0.01 and nothing otherwise. So Question 4' results from Question 3 by applying a *common ratio* – the later decision theorists' expression –, which is chosen so as to revert the choice, and thus plays the same disturbing rôle as the cleverly chosen common consequences in the Allais paradox.

Formally, common consequence and common ratio examples complement each other nicely, in that one group contradicts the statement that for given a and all r, r' ,

$$ap + (1-a)r \text{ R } aq + (1-a)r \text{ if and only if } ap + (1-a)r' \text{ R } aq + (1-a)r',$$

and the other contradicts the statement that for given r and all a, a' ,

$$ap + (1-a)r \text{ R } aq + (1-a)r \text{ if and only if } a'p + (1-a')r \text{ R } a'q + (1-a')r.$$

Allais actually includes certainty as an option in the first pair, but this is really a particular case of the more general patterns exposed by these formulas.

MacCrimmon and Larsson (1979, p. 354-359) also subjected the common ratio example to a parametric analysis, and varying the numerical values, found a rate of paradoxical choices reaching high values, although this required that lotteries promised large monetary outcomes. As before, they connected the brute responses with possible ratiocinations in terms of rules, and paid special attention to consistent subjects. Again two kinds of individuals were found in this group, implicit EU maximizers and paradoxical choosers, though not in a very large number for the last subgroup; however, the very fact that it is non-empty counts as evidence on the above interpretation of Allais's claims⁵⁰.

All in all, the common consequence and common ratio *effects*, as they are now classified, have been replicated a fair amount of times. The most telling results are perhaps those of Kahneman and Tversky (1979), who found violation rates in the region of 50% to 66% for both, when they asked questions with realistic monetary and non-monetary outcomes. The numerical figures in their two-pair experiments appear to be chosen even more efficiently than in earlier ones, supposing that the aim was to refute VNM theory *as an empirical generality*, but we have argued that MacCrimmon and Larsson responded to Allais's normative inspiration and thus did not conceive of a refutation exactly in this way.

4. 4 *An alleged crucial experiment: the utility evaluation effect*

Another experimental scheme raised to popularity belatedly, despite having been suggested by Allais as early as the 1952 Paris conference. The uniqueness part of the VNM representation theorem severely constrains the functions $u(x)$ entering the EU representation obtained in the existence part. Once a zero point and an unit interval have been fixed for $u(x)$, this function becomes absolutely unique (whence the textbook comparison of the VNM index with a temperature scale). In particular, supposing that an individual satisfies the VNM axioms, *the same function* should result if two procedures based on the same zero point and unit interval are applied in succession to construct $u(x)$ from overt choices. A discrepancy between the resulting functions is evidence of the so-called *utility evaluation effect*. The test is crisp, and when proposing it, Allais went as far as to claim that it was “crucial” (p. 247 of the 1953 proceedings).

Allais himself investigated it experimentally by considering two procedures in turn. One – which is now said of the *certainty-equivalence* type – consists in revealing the sure money outcomes that the subject treats as being equivalent to given lotteries, and the other – now said of the so-called *probability-equivalence* type – goes in the opposite direction, thus revealing the lotteries that the subject treats as being equivalent to given sure money outcomes. Allais ended up with a most damaging conclusion for VNM theory, all of his 16 subjects exhibiting the utility evaluation effect (the report is in Allais, 1979, p. 612-613, p. 634). Karmakar (1974) and McCord and de Neuville (1983) investigated the effect afresh with equally clear-cut results.

The fact that this effect is without exception, unlike the consequence and common ratio effects, makes it crucial in the loose sense, which we have found to be used by Popper, of an empirical result that refutes some theoretical compound indisputably. But it does not make it crucial in Duhem's sense of discriminating either between two theories, or between two component parts of the same theory. Quite to the contrary, the sharp conclusion against VNM theory *as a whole* is paid a large price in terms of confusion of possible sources of falsity *within* that theory. Logically, the candidates for violation are not only the VNM independence axiom (A3) and the reduction of compound lottery axiom, as in the previous effects, but also the continuity conditions – related to (A2) – that underlie the way of constructing $u(x)$. To illustrate the point, take a procedure of the above described certainty-equivalence type: it would come to a halt if for some lottery presented to the subject, his answers pointed towards a sure equivalent higher than the best outcome provided by this lottery. This would be a violation of continuity formally identical to that of Marschak's rock climber in section 3. One may rejoin that the utility evaluation effect *itself* would vanish for such a subject, but this would be tantamount to saying that the effect is not universal after all. Also, the construction of $u(x)$ hinges on inferring an indifference statement when the strict preference is reversed – e.g., if the subject claims to prefer 100 euros to a lottery l and l to 99 euros, he is supposed to be indifferent between l and some amount between 99 and 100 euros, and this aspect of the design appears to involve part of the ordering axiom (A1) in the interpretation of the experiment.

Neglectful of these logical possibilities, Allais and most writers after him have treated the utility evaluation effect as another violation of (A3), on a par with the common consequence and common ratio effects. Once again, Allais resorted to the argument that complementarity of chances becomes irresistible to the prudent when these chances add up to certainty. To illustrate, consider two procedures of the certainty-equivalence type that differ by the first lottery they present to the subject. One may intuitively expect that they will entail distinct $u(x)$ *if the lottery is close to certainty for the first, and genuinely uncertain for the second*⁵¹. Now, since procedures of the probability-equivalence type depend on offering sure outcomes, not lotteries, nothing in the argument lets one expect such a strong dependence on the first stage, so it eventually suggests that the two types will be at variance. Here as with the other effects,

the experimentalists' work has filled Allais's theoretical guesswork with substantial evidence⁵².

4.5 *The experiments in the light of alternative theorizing*

The late 1970's and early 1980's witnessed not only the counterevidence to VNM theory that we have gone through, as well as some more that we have not, but also theoretical construals that aimed at accommodating these data, and for the more ambitious ones, at founding the theory of risky choice on a novel basis. By the end of our period, one of the contributors, Fishburn (1988), could mention about a dozen of purported alternatives to VNM theory beside his own. Not all of them are relevant to the present study, which is only concerned with investigating the sense in which refutations occurred, and Duhemian underdetermination was overcome, with respect to this theory. Thus, with due apologies to some of the inventors, we drastically restricted the sample of this paper to four proposals, i.e., Allais's theoretical work up to 1979, Kahneman and Tversky's (1979) "prospect theory", Machina's (1982) "generalized expected utility theory" (GEUT), and "rank-dependent expected utility theory" (RDEUT), which can be traced back to several contributions of the 1980's, the earliest one being Quiggin's (1982).⁵³ The first two contributions are unavoidable, because, as their dates and authors suggest, they might have helped to *shape* the counterevidence against VNM theory. By contrast, the third and fourth contributions are not coincidental with, but reflective upon, the production of counterevidence. They have been included here because each is representative of a major trend of revision – GEUT offering something like a conservative refinement of VNM theory, while RDEUT breaks away from it in significant respects. Some of the received answers to Duhem's underdetermination problem involve attributing a rôle to superseding theories, so there is a philosophical reason for extending our inquiry somewhat in this direction.

Further clarifications are in order before we proceed. First, with typical looseness of terminology, the purported alternatives to VNM theory have been referred to uniformly as "theories" although quite a few of them consisted of a rather thin package – a single mathematical formula, plus some informal motivation and fairly scattered evidence. A philosophical (and indeed any careful) treatment should attend to the distinction *between a specific hypothesis and a full-fledged theory*, so that we will in this case

refrain from using the vocabulary of the field, despite the inconvenience. Second, there are many ways in which a positive alternative – however sketchy – can relate to a refuting test. Ex ante, it can provide it with an abstract scheme, as Allais’s argument about complementarity of chances illustrated. Ex post, it can offer a heuristic account of the results, while perhaps connecting them with those of previous tests – we have also seen this at work with Allais. The account can possibly reach the level of a proper explanation, but we have met no evidence of this achievement at this stage.

With these warnings in mind, let us briefly examine Allais’s “positive theory of choices involving risk”. It actually consists of two distinct generalizations of VNM theory. The first by order of time precedence – it dates back to 1952 – replaces the EU formula with another weighted sum of utility values, in which the weights are *functions of all parameters under consideration*. Formally, if $l = ((p_1, x_1), \dots, (p_m, x_m))$,

$$V(l) = a_1 u(x_1) + \dots + a_m u(x_m), \text{ with } a_i = a_i(p_1, x_1, \dots, p_m, x_m), i = 1, \dots, m^{54}.$$

This new formula is not quite vacuous because of the separation between utilities and weights, but it is badly in need of specification. Allais did not flesh it out until 1988, when he rediscovered RDEUT, to be reviewed below. As early as 1953, he offered a second line of analysis, which prevails in his overall work. It proceeds from the claim that *the dispersion of utility values matters no less than their average value*, which the EUT formula only takes into account. Accordingly, the suitable generalization is either

$$V(l) = g(F(u(x_1), \dots, u(x_n))),$$

where $F(u(x_1), \dots, u(x_n))$ is the probability distribution of the possible utility values, or

$$V(l) = h(Eu(x), \text{Var } u(x), \dots),$$

where the second argument of h is the variance of the utility values, and the unspecified arguments represent as many higher-order statistical moments as one may wish to add.

The last two formulas had in fact spontaneously emerged during the Paris conference, in connection with both the VNM and Savage theories. They were obvious to the other mathematically-minded participants, and if Allais parted company with them, this only because he interpreted them differently. The formulas were indeed equivocal, since they denied EUT the status of an exact theory, but allowed for the possibility that it would be a decent first-order approximation. This last stance was taken by Frisch when he summarized the conference debates:

We must try to proceed further... by looking at... *approximations*... Take any choice structure, say the structure of Paul. In reality, this structure will most likely not satisfy the neo-Bernoullian axioms exactly, but it might do so approximately (p. 253-254 of the proceedings).

Allais emphatically rejected such a soothing view of the performances of EUT. One of his arguments appears to be *a priori* in the strict philosophical sense. The dispersion of utility values is allegedly “the specific element of the psychology of risk” – a necessary part of any risk-attitude, to paraphrase. It follows that a first-order approximation of the V function cannot be correct in general; only a second-order one can conceivably be⁵⁵. The Allais paradox comes to the rescue of the *a priori* reasoning. Here is indeed a thought experiment to suggest that the first-order approximation fails miserably, and by contrast, that the generalizing formulas work.

Given that Allais puts so much weight on the psychological complementarity of chances, one would expect him to represent it in the present framework, but at this juncture, he is distressingly vague. Clearly, this argument requires that $V(l)$ be non-linear, which the equations permit, but the point is to have the equations *entail* it, and for that, the shapes of g and h must be specified. When this is done, one will also be able to answer two important questions left pending, i.e., for what kind of choices the first-order approximation fails – since from Allais's own admission, it does not always do – and whether the second-order approximation is sufficient in general, or in turns calls for refinements. Only in 1979 (see p. 482 and Appendix B2) does Allais become more precise, but not to the point of answering these queries fully. He puts forward the still very general formula for g :

$$V(l) = Eu(x) + H(f(u(x) - Eu(x))),$$

where f is the probability (or density) function of utility values, and H an arbitrary function⁵⁶. Going one step further, one of Allais's followers, Hagen (1979), stated h thus:

$$V(l) = Eu(x) + \alpha \text{Var } u(x) + \beta [E(u(x) - Eu(x))^3],$$

with $\alpha < 0$ and $\beta > 0$. As the sign of the coefficient indicates, the individual is supposed to be averse to the dispersion, but this aversion is asymmetrical, since it depends on how the good outcomes (those above $Eu(x)$) compare with the bad ones (those below $Eu(x)$). At long last, Hagen's formula answers the queries⁵⁷.

With this formula, the Allaisian school might have reached the appropriate degree of theoretical specification, but it would still be open to another criticism: it does not

have theoretical foundations in the same sense as its EU competitors. Allais often claimed to have provided “axioms”, but he just meant by that unifying assumptions from which diverse consequences follow. His work falls short of the standard of axiomatics set up by the followers of von Neumann and Morgenstern, and above all, Savage, which may explain why the latter objected to him that he had no real alternative to offer. Allais's “axioms” involve the utility function $u(x)$ as a datum, whereas section 3 explained that these writers made a point of deriving it. Remarkably, this proves to be not only a theoretical weakness, but *a problem for the experimenter*. If one estimates Hagen’s equation directly, one is left with too many free variables for the same set of choice data, and the experimentation risks becoming a curve-fitting exercise (see Camerer, 1995, p. 627). The advantage of axiomatizations à la VNM-Savage transpires at this point, since they analyze the utility representation into more elementary, qualitative properties that are amenable to a test without the problematic phase of estimating $u(x)$.

In sum, Allais’s “positive theory” is far from constituting the alternative to EUT that he and his disciples claim it to be, and for a long time, it has not even been a theory at all, but rather a heuristic to find one. Even in its final form, it is not clear whether it *explains* the empirical evidence or is but a redescription of the latter. If it matters at all, this is because it contributed to suggest the famous paradoxes, over and above the informal arguments that might have been sufficient for the task.

The second work in our sample, Kahneman and Tversky’s (1979) “prospect theory” briefly emerged as a plausible alternative to VNM theory until some of its weaknesses became apparent. It develops in two parts corresponding to two successive stages of an idealized choice process, which are called *editing* and *evaluation* respectively. An objectively posed choice problem is restated subjectively at the first stage and resolved in this adapted form at the second stage. Kahneman and Tversky's paper is justifiably a classic, but a good deal of its novelty lies in its emphasis on editing, and this part virtually eschews comparison with standard decision-theoretic reasoning. Those concerned with improving on VNM theory have generally focused on the evaluation part, which follows more or less classical lines. Here, the two psychologists propose replacing the probabilities p_i by weights of the simplest possible functional form:

$$V(l) = a(p_1)u(x_1) + \dots + a(p_m)u(x_m).$$

The outcomes are monetary, and for simplicity, the authors focus on lotteries having at most two non-zero such outcomes. Some experimentalists of the 1950's and 1960's, like Edwards, had proposed the same formula in response to the first apparent violations of VNM theory, generally interpreting $a(p)$ as a *subjective distortion* of the given probability vector p ⁵⁸.

Kahneman and Tversky rejuvenate this line by making new hypotheses of about the u and a functions. As regards the former, they conjecture that $u(x)$ is defined on *changes* in wealth rather than total wealth, and that this function exhibits risk-aversion (concavity) on the range of gains and risk-love (convexity) on the range of losses. Since 1979, empirical support has built up in favour of this conjecture, which has in effect displaced Friedman and Savage's (1948) failed attempt at determining the curvature of u . As regards the latter function, Kahneman and Tversky begin by assuming that a is increasing between the extreme values $a(0) = 0$ and $a(1) = 1$, and then propose substantial restrictions, *each of which is dictated by a well-established effect*, which makes their technical discussion exemplary. One of these restrictions, i.e., that low probabilities are overvalued and high probabilities undervalued, directly connects with the common consequence effect. They also deal with the common ratio effect, and if they do not envisage the utility evaluation effect, this is because it was little known at the time. Having checked for the logical compatibility of the shape restrictions, they eventually suggest to take a to be convex except in the vicinity of 0 and 1. This probability-related analysis has not gained the same acquiescence as the utility analysis of the paper, in part because RDEUT, which defines a differently, has superseded "prospect theory" in most specialists' opinion, in part because those who have adopted the same definition have sometimes favoured other shapes⁵⁹.

The previous formula for $V(I)$ permits violating the dominance principle, and this is the main reason why the "alternative account" that Kahneman and Tversky's (1979, p. 263) announced did not really take off. Once this possibility was demonstrated, most decision theorists lost interest, because they strongly believed that a proper alternative to VNM theory should retain the principle (in this respect, Allais is representative of the community as a whole). Kahneman and Tversky (1979, p. 283) had actually guessed that a non-linear a conflicted with dominance preference, but they had responded by making the editing stage responsible for the detection and discarding of dominated lotteries (ib., p. 275). This is an extraordinary move if one thinks of it, since most

lotteries are dominated by others in a standard VNM setting, so that the evaluation stage would be left with near to nothing to operate on.

Independently of the conundrum created by dominance, those accustomed to EUT axiomatizations could complain against “prospect theory” on the formal grounds that it lacked foundations. The evaluative part revolved around a single formula with utility as a datum, a formula which enjoyed some inductive, but no deductive support, and was not axiomatized. In fact, Kahneman and Tversky had proposed a specific hypothesis rather than a full-fledged theory, which in part reflected the different working habits of decision psychologists and mathematical decision theorists.

In retrospect, leaving aside the new definition of u , “prospect theory” matters because it sharpened and helped to reorganize the empirical counterevidence of the time to VNM theory (though not the utility evaluation effect). Also, it signalled a major shift in the understanding of experiments, which henceforth became directed at refuting or confirming empirical regularities instead of entering into a normative assessment of rules of conduct, as was the case in the work closely following Allais.

The third potential alternative to be considered here, Machina’s (1982) “generalized expected utility theory” (GEUT) had a twofold aim, for one to weaken VNM independence so as empirically to account for several known effects at once, and for another, to preserve as much as possible of the apparatus developed by VNM theorists to deal with risk-attitudes. The strategy was to strike a balance between conservatism and revision, as it were. In terms of the axiomatization of section 3, it consists in keeping (A1) and (A2), while replacing (A3) by two conditions, (H1) and (H2), that are directly stated in terms of the V function.

(H1) The function V representing \leq over L is differentiable in the probabilities.

Barring technicalities, it is clear what this condition means to achieve: while linearity in the probabilities is a global and exact property, replacing it with differentiability will preserve it *locally and as an approximation*. Frisch and others had pointed out this way out, and Machina is now following it to the end. A precise formulation of (H1) requires mathematical care because of the non-standard domain L^{60} , and Machina simplifies his task by assuming that the set of outcomes X comprises of all possible numbers between 0 and some positive M . This is sufficient if the aim is to handle experiments with money outcomes as well as the more standard economic applications. It follows from (H1) that for any chosen reference lottery l , there exists a

function $u_l(x)$ entering a linear formula that approximates V around l . Thus, something of the VNM representation theorem can be salvaged, as one could hope. Further – a most useful contribution – the VNM analysis of *preference for dominance* and *risk attitudes* turns out to be essentially preserved. The first step is to check is that these objects can be investigated locally as they were globally, i.e., in terms of the monotonicity, concavity and convexity of the u_l . Then comes a second, more ambitious step, which is to reconstruct global properties in terms of local ones, and then show that some of the existing theorems obtained under VNM theory still hold for GEUT⁶¹.

Besides its applications to economics proper, for instance to insurance theory, the preceding analysis permits handling the experimental effects of this section, as well as some more. By suitably adjusting the curvatures of the u_l , one can make V compatible with choices that are sometimes cautious, and sometimes not, as in the common consequence or common ratio effects. But clearly, it would be too easy a game to fix the local utility functions just as the subjects' answers require, and this is where the supplementary condition (H2) comes in. It takes into account a *systematic* property of the observed violations that we have not yet brought out. In Allais's Questions 1 and 2, x_1 dominates x_2 , and y_1 dominates y_2 , in the sense relevant to the dominance principle – that is, x_1 (y_1) gives more weight than x_2 (resp. y_2) to the more valuable outcomes. Other experiments performed on the common consequence effect obey this pattern, which can be formalized thus: in the sense made precise by the concavity of his u_l functions, the individual becomes *less cautious* when he moves from a choice between lotteries x_1 and y_1 to a choice between x_2 and y_2 such that x_1 dominates y_1 and x_2 dominates y_2 . As Machina (1983) explains, the same pattern of variation occurs with the common ratio and the utility evaluation effect. This finding motivates the next condition.

(H2) If a lottery l dominates another lottery l' , u_l will exhibit no smaller risk-aversion than $u_{l'}$.

In the particular case of three monetary outcomes, (H1) and (H2) admit of an elegant two-dimensional geometric representation⁶². The individual's indifference loci are not parallel line segments anymore, as VNM theory requires, but continuous and smooth curves (because of (H1)), which “fan out”, i.e. become steeper in the direction of increasing dominance (in connection with (H2)). Given that (H1) makes such a wide generalization step, (H2) is essential to the determinateness of GEUT. Without it, one

can doubt that this theory would explain anything at all. But even with it, does it really explain the paradoxes? Machina claims that “all follow from a single assumption [= (H2)]... which leads to further refutable restrictions on behavior” (1983, p. 282). This is correct only if one does not take “follow” in a straight logical sense. For (H2) does not tell *by how much* the curves become steeper, and lacking quantitative precision, it cannot logically entail, for instance, that an individual satisfying it will answer Allais’s questions in the paradoxical way. In fact, as the formal statements indicate, (H1) and (H2) together are compatible with the individual’s choosing in the VNM way. The sense in which GEUT *explains* the paradoxes must then be qualified thus: for any given pair of choices that violates VNM theory, if one assumes (H1), one can find a quantitative specification of (H2) that entails the violation. This sense of “explanation” involves an element of curve-fitting⁶³.

GEUT attracted, and still attracts, considerable interest from decision theorists, but few would describe it as a proper alternative to VNM theory, and the reasons for this appear to be twofold. For one, despite its axiomatic flavour, it does not consist of an axiomatic system in the received sense because it uses $V(I)$ as a primitive term. There are serious, perhaps insuperable, difficulties involved in the rendering of (H1) at the qualitative level of a preference relation. For another, and more importantly, (H2) is only weakly predictive and weakly explanatory, as was just explained. This is not to say that informative tests cannot be performed, because (H2) excludes some VNM configurations (those with “fanning in” or no fanning at all), and can anyhow be rejected by the data despite being weak. The experimental work along these lines has suggested a mixed record⁶⁴. In retrospect, its significance may primarily rest with the generalizing step contained in (H1), which implements the idea that VNM theory can be stated as an approximation.

4.6 Rank-dependent expected utility theory

The last item in our narrow list, RDEUT is by now the best regarded, and it can indeed be argued that it deserves the highest consideration. Historically, it is a curious case of intellectual convergence, because its basic mathematical formula was worked out independently by Quiggin (1982), Yaari (1987), and Allais (1988), and these writers came across it starting from somewhat different motivations. Formally, take a lottery

$$l = ((p_1, x_1), \dots, (p_m, x_m)),$$

with the outcomes x_1, \dots, x_m being ranked in that preference order from the lowest to the highest. Then, the RDEUT evaluation is

$$V(l) = a_1 u(x_1) + \dots + a_m u(x_m),$$

where $a_1 = f(1) - f(p_2 + \dots + p_m)$, \dots , $a_i = f(p_i + \dots + p_m) - f(p_{i+1} + \dots + p_m)$, \dots , $a_m = f(p_m)$, and f is an increasing function from the 0-1 interval to itself that satisfies $f(0) = 0$ and $f(1) = 1$. Equivalently,

$$V(l) = b_1 u(x_1) + \dots + b_i (u(x_i) - u(x_{i-1})) + \dots + b_m (u(x_m) - u(x_{m-1})),$$

with $b_1 = f(1)$, \dots , $b_i = f(p_i + \dots + p_m)$, \dots , $b_m = f(p_m)$.

This formula generalizes that of VNM theory, which is recovered by putting $f(p) = p$. As the first equation makes clear, it consists of a weighted sum of utilities in which the weights a_i are functions of probability *sums*, not of individual probability values as in “prospect theory”. The sums compute the probabilities that the realized outcome will *reach or exceed* the successive preference ranks; for $p_1 + p_2 + \dots + p_m = 1$ is the probability of getting at least the worst outcome x_1 , $p_2 + \dots + p_m$ is the probability of getting at least the second worst outcome x_2 , and so on until one reaches the probability p_m of getting exactly the best outcome x_m . So what f transforms is the *decumulative* distribution (i.e., one minus the cumulative distribution, as usually defined) that is associated with the probability vector p .

The second equation suggests an informal rendering in terms of a sequential evaluation process. Taking each x_i in turn, the individual applies to its value $u(x_i)$ an abatement ratio b_i that depends on how uncertain he is of receiving x_i or a better outcome. He knows that he will enjoy at least the utility of x_1 with certainty, hence $b_1 = 1$, that he will enjoy at least the utility increment of x_2 with some uncertainty (whence $b_2 < 1$), and so on until he considers the utility increment of x_m , which he is maximally uncertain to enjoy (whence $b_m < \dots < b_2 < 1$).

RDEUT has a synthetic value that its inventors perhaps did not realize fully, and in any case did not stress, because they had different points to make in using the same mathematics. First of all, it solves the problem that had plagued “prospect theory”. For the above definition of $V(l)$ satisfies the dominance principle by the increasing property of f , and it is demonstrably unique in satisfying the principle within the class of

functions having a similar weighted sum structure. This uniqueness property was Quiggin's main justification for adopting RDEUT, and it is also a reason for Allais, who had toyed with weighted sums and also argued for the rationality of the dominance principle.

Second, the probability transformation function f comes with an interesting interpretation in terms of *risk-attitudes* that was not so obviously available with other functions in the same class. Quite naturally, the early experimenters tended to think of their $a(p)$ as of "subjective probabilities", in the – non-Savagean – sense of a cognitive distortion of the given lottery probabilities. Thus, overweighting of small probabilities meant that the subject regarded the corresponding events as more probable than they were; and symmetrically for underweighting of large probabilities. With RDEUT, the dominant interpretation – fixed by Yaari and Allais – is that f reflects the individual's propensity to take chances, not his misperception of their magnitude. This semantic line is comforted by mathematical results that connects with the VNM analysis of risk attitudes in terms of the u function. In a sense that these results make precise, f and u are dual to each other⁶⁵. Very roughly speaking, the risk-aversion property that VNM theorists identify by saying "the u function is concave" is now captured by saying "the f function is convex". Quite a few theorists today have replaced one statement by the other, thus dramatically breaking up with the received analysis⁶⁶.

Last but not least, the RDEUT evaluation satisfies a weak form of VNM independence, the so-called comonotonic independence, which is easily interpretable and can be tested exactly to the extent than its ancestor condition can (i.e., with the same possibilities for confusion of possible sources of falsity). Two lotteries l, l' in L are said to be *comonotonic* if $l = ((p_1, x_1), \dots, (p_m, x_m))$, $l' = ((p_1, y_1), \dots, (p_m, y_m))$, and the outcomes $x_1, \dots, x_k, \dots, x_m$ and $y_1, \dots, y_k, \dots, y_m$ are ranked by increasing preference order.

(A3⁻) (Comonotonic Independence) For all l, l' in L , and all numbers a between 0 and 1 (0 excluded), if l and l' are comonotonic and for some k , $x_k = y_k$, then

$$l R l' \text{ if and only if } l^* R l'^*,$$

for all l^* and l'^* , which are identical to l and l' , respectively, except that $x_k^* = y_k^*$ with x_k^* possibly different from x_k .

Compared with (A3), this condition limits the replacement of a *common consequence* to those cases in which the initial lotteries are comonotonic and the

replacement does not upset the order of outcomes in the lotteries. It is easy to check that the Allais paradox choice does not violate $(A3^-)$, which is well suited to handle the common consequence effect generally. The new axiom can be defended on the ground that l and l^* on the one hand, and l' and l'^* on the other, exhibit the same qualitative pattern of risk, so that there seems to be no reason why the preference between l and l' should differ from the preference between l^* and l'^* . This is the gist of a *rationality* argument for obeying $(A3^-)$ and not obeying the stronger $(A3)$ ⁶⁷.

Comonotonic independence can be included in a full-fledged axiomatization of RDEUT, so as to reach the formal standard of VNM theory and make comparisons with it more definite. In the more advanced framework in which probabilities are not given, there exists a related condition, which is the comonotonic counterpart of a classic axiom in Savage's system, and this condition can similarly be embodied in a full-fledged axiomatization. Actually, the move on this front was initiated by Schmeidler (1986) independently of the move against VNM theory⁶⁸. We have excluded Savage from our investigation, and must then gloss over an important development here, but it needs stressing because it is part of the collective discovery of rank-dependent evaluations, and conceptually, further proof of their theoretical significance.

From the viewpoint of empirical performance, RDEUT is open to the same problems as GEUT. The curve-fitting element is obvious from the solutions given by Quiggin and others to common consequence or common ratio problems. In general, given an observed choice in a two-pair experiment, concavity or convexity restrictions on both u and f are not sufficient to entail either this choice or its absence, and one should take the step of fixing some numerical values for both functions. The comparative empirical record is mildly favourable to REUDT⁶⁹, but it seems clear that this has played only a minute rôle in tilting the decision theorists' scales towards it - the main reasons being the convergence of arguments to support the weighted sum formula and the availability of an axiom system which articulates them coherently. The success of RDEUT is relevant to Duhem's underdetermination problem only on the view that the solution to this problem requires that a superseding theory be available, but more generally, it tells us much about the way decision theory evolves and perhaps makes progress.

5. A QUASI-DUHEMIAN ACCOUNT OF THE HISTORICAL SEQUENCE

5.1 *Counterexamples, anomalies, and effects*

Right from its beginning, VNM theory was surrounded with *counterexamples*, a situation which its exponents were generally willing to acknowledge. The troubling cases were suggested by casual observation of human life and required ingenuity only to connect them with particular axiomatic conditions (such as continuity in adventurous behaviour and reduction of compound lottery in multi-stage gambling). Those who accepted and even provided the counterexamples conceived of them not as *refutations* of the theory, but as *directions to specify its domain of application* more precisely than the founders had done⁷⁰. A common idea was to specialize the theory in economic applications, but this turned out to be a red herring. Attempts were made to revive the economists' exclusion of the pleasure of gambling, and in the final one, this led to a division of risk attitudes, some but not all of them falling within the competence of the theory. What is noteworthy here is that VNM theorists could not assign a domain without resorting to the rationality concept, for which they had no independent definition, and which they explored *by using the concepts of the theory itself*, a circular move that was perhaps unavoidable.

Anomalies are another particular case of refutations, bordering refutations, but not identical with them. They do not impact on any statements, but only generalities, and they enjoy the dialectical privilege of exceptions, which both infirm the rule and confirm it, provided they are neither too deep nor too numerous. Thus, anomalies can be seen as refutations in a virtual state that is realized only if they are reconceptualized or widely replicated or both. Since Kuhn and Lakatos, it has become standard to claim that scientific theories are accompanied with anomalies. However, the first counterexamples to VNM theory were *not even treated as such*, the first case being arguably the Allais paradox, which does not belong to the initial stock. Lakatos (1970, p. 120, n.2) famously claims that "in actual history new theories are born refuted; they inherit many anomalies of the old theory". This does not appear to apply here. One of the reasons is that VNM theory developed in something like a vacuum, because the previous economists had eschewed any serious investigation of risk attitudes.

The fact that there was so little done on this major topic before *Theory of Games and Economic Behavior* is the strongest argument in favour of the one-sided strategies

of the early years. Had decision theorists yielded to the fascination of counterexamples, microeconomic textbooks would look very different today. Insurance economics, a good deal of theoretical finance, as well as many of the received game-theoretic applications to industrial organization, employment contracts, investment policies, or more recently auctions, took off thanks to the EU formula and the trick of representing the agents' risk attitudes in terms of the utility function. In order to protect these usable components from criticism, the early theorists defended VNM theory *as a whole*, despite the clear suggestion of the counterexamples that some axioms were more dubious than others. They were holists, but technically not Duhemians, since an analysis of domain restrictions is not to be found in *TP*. They were not Kuhnians or Lakatosians either, because, as just said, they had no use for the concept of an anomaly, to which domain restrictions are typically referred in the growth-of-knowledge literature.

The Allais paradox acquired the status of an empirical refutation without going through the intermediary stage of an anomaly, but this happened for reasons that parallel those which turn an anomaly into an empirical refutation. For one, a deeper analysis uncovered its significance, and for another, it was transformed into an empirical generality of some kind. The paradox became an *effect* (the common consequence effect), a metatheoretical concept of decision theory that may now be analyzed. Put abstractly, it is an empirical phenomenon that is endowed with some regularity, though it admits of exceptions, and is accompanied with some explanation that is both causally relevant and psychologically intelligible. There is often a semantic suggestion that the phenomenon can be reproduced with some success by the experimenter, which would imply that the causal explanation is adequate at least for manipulative purposes. An effect belongs to the outside world of observation, but typically receives its name from the factor that supposedly triggers it out. Mathematical decision theorists like Machina (1983) tend to identify effects with formal statements that enter logical relations with each other and with axiomatic conditions. Kahneman and Tversky (1979), and decision psychologists in general, are much less formal. They do not worry so much about entailment as about compatibility relations that support causal explanations. Also, psychologists often stop their explanations of choices at the level of an effect when it is sufficiently unexceptionable, while mathematical theorists

regard them as being at best low-level generalities in need of a unifying deductive explanation.

Here is how this concept relates to our main theme: once a counterexample such as the Allais paradox is registered as an effect, one major item in the Duhemian compound – i.e., *the observational record* – disappears from consideration. In the particular instance, the registration process obeyed the pattern that Duhem let one expect for a solution to the underdetermination problem in general – there was a relatively lengthy exchange of pros and cons with an overall conclusion that the cons weighed more heavily. Indeed, the anti-EUT continuously brought out challenging facts, somewhat like the undulatory theorists, while the pro-EUT mostly limited themselves to reinterpret the evidence, somewhat like the corpuscular theorists. At the risk of exaggeration, one may liken any of the three effects discussed in this paper to the lower-level generality that was at issue in Foucault's case, i.e., that under certain specifiable and reproducible conditions, light moves more quickly in the air than in water.

To claim that an effect, in the above sense, is reasonably well established presupposes that some check has been made of the possibly disturbing rôle played by the *concrete conditions* in which the experiments took place. This implies the more basic requirement that these conditions be identified with some objective precision. What were the physical devices employed to put the choices to the subjects, and especially, to convey the probability numbers to them? When the experimenter varied the pairs of lotteries, how many did he put in a row, and did he present each pair only once or several times, and in the latter case, after what time lag? Were the chosen lotteries drawn, and if so, were the resulting money outcomes effectively paid for all, some, or none of the drawings? Were the subjects students, businessmen, or ordinary people? Had they been exposed to the concepts of the theory before? When the experiment involved some discussion of the VNM conditions, at what level of abstraction was it conducted? In the period under consideration, the experimental reports were terse, and from what can be inferred, not all of the problematic conditions were identified, let alone controlled for. For instance, experimenters may have worried about the possibility that students reacted differently from ordinary subjects, but they nonetheless continued to use the same abstract questionnaires for the various social groups. They hardly addressed the tangled issue of real payments and what differences it made to introduce them, both in terms of stronger incentives and wealth effects

changing the subject's greediness. These loose ends are easy to point out now that experimental economics circulates more or less standardized lists of factors to check⁷¹. Still, the 1970s and 1980s witnessed a number of replications, with conditions changing *at least from one experiment to another*, and since the aim was to establish effects of a broadly qualitative kind, one may conclude that this unsystematic variety was sufficient. The utility evaluation effect can be said to be well supported, and the common consequence and common ratio effects, to be recognizable from the data, the former probably with a better record than the latter.

5.2 *Confusion of sources of falsity once the empirical record is established*

In Duhemian terms, once the step is taken to regard the observational record as unproblematic, the remaining sources of falsity are the primary hypotheses and *some* of the auxiliary hypotheses (a good deal of them having been handled at the previous stage). In VNM theory, the primary hypotheses are the preference axioms, which were tested by investigating choices in an artificial context of questions and answers. It would have been possible to register their spontaneous risk-taking choices, as Friedman and Savage (1948) coarsely did when they observed that the same people gamble and take out insurance. There is a threatening possibility that the effects, although well evidenced among the subjects of an experiment, do not carry through to the actual, uninfluenced behaviour of nondescript individuals, and are thus irrelevant to the theory since it is *that* kind of behaviour which it means to capture in the end, whatever the precise frontiers of the application domain. Logically, the experiments that establish an effect are conclusive against VNM theory only under the supplementary assumption called *parallelism* or *external validity* in today's discussions of experimental economics. There is a clear contrast with the *internal* validity assumptions that were exemplified in the last subsection. To some extent, parallelism involves a metaphysical claim of the "uniformity of nature" or "limited variety" kind, which is so general as to underlie any experimental work, but it can also be viewed as making an empirical - and even possibly testable - claim on how humans behave under different environments. If the latter line prevails over the former, parallelism should be included in the list of auxiliary hypotheses which impinge on the conclusion that the effects refute VNM theory⁷².

Importantly, work based on this theory has persisted up to now in economics, both at the applied and theoretical level, and this can be explained along various lines. One is the economists' well evidenced stubbornness when they meet facts that challenge their theoretical outlook⁷³. But a perhaps more attractive explanation is that they are reluctant to accept parallelism in its empirical interpretation. For they are not impervious to *any* choice evidence, but only to that which does not take certain standardized forms. In particular, choices should be made in the context of a genuine economic activity, most typically one involving significant money flows, and they should be observed from the outside rather than reported in verbal statements. Some, though not all, economists would add that choices should be observed across a sufficiently large number of different individuals for the most idiosyncratic features to cancel out⁷⁴. There is still another interpretation, which is exclusively applicable to the topic of this paper, i.e., that the reluctant economists take the view that VNM theory is inaccurate rather than plainly false, and should be preserved as an approximation for lack of a better alternative. This view was floated as early as the 1952 Paris conference, and we have seen that it connects with Machina's "generalized expected utility theory" (GEUT). Even if Machina has shown that some of the standard economic theorems still hold in his framework, other applications require the linearity of the EU formula exactly, not just approximately. Being more uniform, the previous suggestion that economists reject parallelism is perhaps the best suited to redeeming their work.

Both experimental and non-experimental applications raise the little noticed problem that the subject's *choices* are imperfectly related to his *preferences*, which are the true object of the axiomatic conditions. Identical in this respect, decision theorists and economists work with such an impoverished semantics that they ignore this possible discrepancy, but the experimentalists often rediscover it obliquely. For instance, some have worried that subjects may take a strategic attitude towards them – play against them as it were – because of the contrived experimental conditions, and thus may *not* express their preferences genuinely⁷⁵. If this kind of attitude is to be excluded, another parallelism assumption must be added to the first, and to decide whether or not it counts as an auxiliary hypothesis is again a matter of striking a balance between metaphysical and empirical claims.

The many reasons for underdetermination did not stop decision theorists (and at least some economists) from judging that VNM theory was empirically refuted,

however differently they phrased this conclusion. Was it based on a two-sided comparison of the effects with the target theory, or did it appeal to alternative theoretical hypotheses? Here we meet an issue that attracted the attention of the philosophers of science when Lakatos's "methodology of scientific research programmes" (MSRP) made the existence of these hypotheses essential to the meaning of falsification. One of his claims is that a theory T counts as falsified *only if* there already exists an empirically progressive alternative T' within the same research programme. Lakatos makes this necessary condition *also sufficient*, and thus allows for the possibility that T be falsified simply because there exists a superior T' , i.e., even if there is no counterevidence⁷⁶. Such a wedge between the senses of "falsification" and "refutation", the better understood term, makes the sufficiency claim dubious, so we leave it out and just discuss the more basic necessity claim.

The time order matters here, as it should in an historicized conception like the MSRP. Granting that VNM theory was generally taken to be refuted by the early 1980s, the natural candidates to the status of a superseding alternative theory are twofold, i.e., Allais's long preexisting "positive theory" and Kahneman and Tversky's 1979 "prospect theory", and we have shown that both fall short of the desired status. We do not mean to say, of course, that unorthodox theoretical construals did not underlay the counterevidence, in the senses of suggesting it and helping to conceptualize it; but Lakatos's claim is more exacting than this point about theory-laddeness. As also explained, GEUT and "rank-dependent utility theory" (RDEUT) do meet the requirement for being theories, but only the latter is clearly an alternative, and historically they came after the fact. In sum, VNM theory exemplifies the case of a T that was taken to be refuted *without an alternative T' being yet in place*. This is a clear rebuttal of Lakatos's claim in its descriptive reading, and to say the least, a challenge to its normative reading, given the proclaimed endorsement of the "scientific elite" by the MSRP. By contrast, the conclusion is reassuring both for Duhem and Popper, who do not condition the occurrence of a refutation on Lakatos's heavy condition⁷⁷.

The time order is not all that matters in this discussion because the MSRP criterion for T' to count as a *superior* alternative would anyway not be met easily. This criterion requires in particular that T' have "excess empirical content" with respect to T , i.e., "predict novel facts", but that "all the unrefuted content of T [be] included", so that " T' explains the previous success of T " (Lakatos, 1970, p. 116)⁷⁸. Now, GEUT is

mathematically stated as a logical weakening of VNM theory, which means that it cannot make predictions that the latter would not already make. It is only in the *psychological* sense that it has brought novel facts to attention – in principle, they could have been derived from the earlier theory⁷⁹. RDEUT is in a somewhat different situation. It is similarly a weakening of EUT if one allows for the probability-transformation function to be the identity, but it is not anymore so with the commonly envisaged non-linear shapes, and it can then deliver new predictions compared with VNM theory. However, with RDEUT understood this way, the other Lakatosian condition that T' recovers *all* unrefuted predictions of T becomes problematic. To summarize bluntly, the genuine alternative theories appear to be *sometimes too weak and sometimes too strong*.

Although progress is not the main topic of this paper, this finding should be related to the definition of this concept not only in Lakatos (1970, p. 118-119), but also in Popper (1963-1972, p. 240-243). The former took from the latter the grand view that science progresses by offering more informative theories that do not lose in “corroborated” content – the best of two worlds as it were. Duhem is too sceptical to make the heavy demands on progress of the two falsificationists, and his philosophy of science seems to be compatible with the view that RDEUT constitutes a genuine advance of decision theory with respect to VNM theory⁸⁰.

5.3. Duhemian confusion of the sources of falsity within the theory

As we argued, Duhem's underdetermination problem can sometimes be resolved to the point where the scientist holds a primary hypothesis responsible for the refutation, whence the question: can decision theorists say what exactly was refuted in VNM theory? The utility evaluation effect is telling against VNM theory in general, but involves too much of its content to be relevant at this stage. More to the point, the common consequence and common ratio effects are motivated by an attempt at questioning the VNM independence axiom (A3) specifically, and what they do achieve well is to *keep out of consideration the ordering and continuity axioms (A1) and (A2)*. Philosophers of science should praise those theorists, like Allais, who conceived of the discriminating experiments, and those experimentalists, like MacCrimmon and Larsson, or Kahneman and Tversky, who implemented them with relatively clear-cut results.

Popper's recommended procedure – i.e., once the theory is decomposed axiomatically, to devise tests for its axioms individually – appears to have been carried out with some success in the particular instance.

A comparison with earlier work will help to appreciate this success more fully. Mosteller and Noguee (1951) attempted to test VNM theory by using the strategy of estimating individual utility functions first. More precisely, they adapted the certainty-equivalence method of estimation, which was described in 4.4, because they doubted that the individuals satisfied the transitivity condition in (A1) exactly. So they put the same questions several times over to their subjects and defined the monetary equivalent of a lottery for one of these subjects by some statistical criterion. Similarly, when they moved from the estimation to the testing stage, they repeated their questions and stated their results in terms of another statistic. Mosteller and Noguee's conclude that "the notion that people behave in such a way as to maximize their expected utility is not unreasonable" (1951, p. 403), but even such a moderately favourable conclusion is far-fetched in view of the statistical assumptions that predetermine their data. Although this may be excused by the date, their work appears to be a sad example of mismanagement of Duhem's underdetermination problem. They *increased* the range of underdetermination instead of reducing it. Conceptually, their mistake was to involve (A1) in the analysis of the experiment, and it was aggravated by their way of replacing it not with a proper axiom, but with a compound of auxiliary hypotheses.

The generous interpretation of early experimental pieces like Mosteller and Noguee's is that they did not really aim at checking the axioms of VNM theory, being primarily concerned with lower-level propositions that this theory made it possible to formulate, such as those relative to risk attitudes or the subjective perception of probabilities. With this interpretation, it would become possible to claim that decision theorists were competent scientists throughout. In the first years, they would not have bungled the underdetermination problem, as we just suggested, but rather ignored it because they were working *within* the theory, and not yet curious about its axiomatic basis. The succession of a dogmatic and critical stage is standard in growth-of-knowledge accounts of science, like Kuhn's and Lakatos's, and although an oversimplification of our narrative, it appears to capture a grain of truth in it.

However we conclude on the first period, we must temper the praise of the second, because *it did not resolve the underdetermination problem fully*. The common

consequence and common ratio effects hit (A3) and the reduction of compound lottery axiom jointly. Critics of VNM theory usually assumed that it was the former, not the latter, which was responsible for the effects, but one is struck by the meagre evidence they had at their disposal for this strong conclusion. Essentially, there was the point claimed by Allais, and to some extent borne out by experiments, that paradoxical choices were about as frequent among subjects who knew the probability calculus as among subjects who did not. This is hardly impressive because, as we mentioned, reduction can fail for more than one reason. Beside not knowing the multiplication rule of probabilities, an individual can make mistakes in applying it, and it is also conceivable that he masters the rule fully but doubts that probabilistic independence applies to the multi-stage lottery presented to him. Here, decision theorists are found lacking *on Duhemian terms*, which shows that the flexibility of the account is not without limit. There is an impressive convergence of the three effects, but it does not point towards (A3) alone, and more experiments were actually required.

It is indeed easy to test (A3) alone – select two pairs of lotteries with a common consequence that is not the same in each pair, regardless of the fact that some of the lotteries may be compound. Outside our period of study, this straightforward test was performed with the striking result that the proportion of violations was significantly weaker than that of the common consequence effect, and these violations proved to be irregular, arguably lacking the systematic element of the latter effect (see Conslík, 1989). An interesting view that has recently emerged is that the systematic violations of VNM theory must be referred to *independence in reduced form*, which is a joint implication of independence and reduction⁸¹.

Pending cross-confirmation of this hypothesis, we can say no more on it, but a point of philosophical interest has hopefully appeared in full light. *The solution to Duhem's underdetermination problem is sensitive to the axiomatic decomposition adopted for the theory*. If the responsibility of the decision-theoretic effects eventually lies with a condition that is neither VNM independence, nor reduction of compound lotteries, the standard axiom system becomes unhelpful, and decision theorists had better replace it by a system – if it can be devised logically – that includes the culprit as one of its axioms. In terms of the standard system, there is only one sensible answer to underdetermination, which is – disappointingly – that independence and reduction are *jointly* responsible for the refutation. To the best of our knowledge, those who offer the

axiomatic method as a remedy to underdetermination have not pointed out that the system can limit the degree to which the latter can be overcome.

5.4 *The normative discussion in the quasi-Duhemian account*

As they are usually understood, anomalies and refutations belong to the class of empirical counterexamples, the only ones thoroughly discussed in philosophy of science. But the VNM axioms were thought of as putative rationality conditions, and the theory was assessed in terms of normative force no less frequently than it was in terms of empirical performance. Sometimes the two viewpoints combined subtly, as in the Allaisian experiments about the wise men's rules of choice. A major question for our Duhemian perspective is whether it can make room for both the normative and empirical strands of the discussion.

The received history of VNM theory answers the problem in reorganizing the narrative roughly thus. Allegedly, the theory was offered as being primarily empirical. But then came Allais with his paradoxes, which made it clear that it was descriptively false. The tenors of EUT had to recognize the fact, especially Savage, whose own answers to Allais's questionnaire documented the failure, and what they did afterwards was to *retreat to the normative*. Soon followed by others, Savage, de Finetti, Marschak and Samuelson buttressed VNM theory by expanding on its normative justifications. They left the empirical ground to the experimentalists, who developed the new “theories” on this ground more or less exclusively.

The historical part of the paper contradicts this reconstruction in a number of ways. First, VNM theory was proposed as a theory of rational choice right from the beginning – this is how it appears in *Theory of Games and Individual Behaviour* as well as in Marschak (1950). Second, Allais's paradoxes were thought experiments specially devised to challenge the claim that VNM theory was normatively compelling, and nobody took them to constitute empirical refutations until they were transformed by experimentalists much later. Third, even at this late stage, defenders of VNM theory had empirical arguments at their disposal and sometimes made use of them. The only indisputable claim is that Samuelson and Savage neglected these arguments, but they had hardly paid attention to the empirical side of the theory at first⁸².

In sum, the two components were present at all temporal stages of the discussion, and often, though admittedly not always, in the one and the same piece of work. This suggests focusing on *arguments*, rather than times or individuals, as the proper units of analysis. Here is a very brief review of those which can be expected to link the normative and the empirical with each other.

It is at least arguable that *a rational rule of conduct stands a better chance of being adopted than an irrational one*. There is some evidence that the VNM theorists conceived of this linkage, and by this plausible attribution, one makes sense of the claim that the normative defences – especially, those mentioned in section 3 - influenced the acceptance of von Neumann and Morgenstern's theory as an empirical one.⁸³ The linkage of rationality with frequency can also be presented in the converse way: gross violations of rationality are not met very often. This version surfaced in some of the controversies about the paradoxes, and up to now, it remains the best reason that decision theorists have at their disposal to impose the dominance principle on their empirical theories.

Another connecting claim is that *strict rationality conditions give structure to the choice data*, whether these conditions are themselves put to the test (like VNM independence) or assumed to hold in order to facilitate the test of other conditions (like transitivity). This linkage is unproblematic, contrary to the previous one, but it may be unspecific to the theories of rational decision making, since it is always the case that a hypothesis easier to test the logically stronger it is.

Still another connection is that *the normative force of a rationality condition may be subjected to a test*, as in the argument about Allais's paradox. Remember that Allais was initially concerned with experimentation only to check that rational and prudent people made the choice that he had foreshadowed in his thought experiment. He avoided crude circularity by assuming that rationality and prudence are qualities that can be recognized in someone by common consent. Savage also relied on an empirical test of normativity, though a different one. He argued in effect that if a man decides in a rational moment not to comply with a rationality condition, this establishes that the condition has little or no force after all. He avoided crude circularity by assuming that a rational moment can be recognized at the time and depth of thinking given to the issue.

We finally mention the connection implied by *the cognitive preconditions of rational decision-making*. Since Simon's work on bounded rationality, it is a well taken

point that in order to have any normative force, a rule of decision must make feasible demands on the individual's ability to collect information and make computations. This new linkage of the normative and the empirical is but a contextual way of making good the meta-ethical principle that “ought” implies “can”. We found it mentioned by MacCrimmon and Larsson at the expense of VNM theory, but it can be no less damaging against alternatives.

How does this sketchy and no doubt incomplete list reflect on the quasi-Duhemian account? Without providing the full argument, we submit that it does not undermine it, but rather serves to determine it further. Some of the suggested links between the normative and the empirical are expressed in claims that can *themselves* be tested, whether directly or indirectly, thus delivering a new range of possibilities for Duhemian underdetermination. One such claim is that gross violations of rationality cannot be met very often. A possible check is to observe whether or not actual decisions obey the dominance principle, a somewhat neglected test. If the answer is significantly negative, there will be a – Duhemian – choice between giving up the claim or disconnecting the dominance principle from rationality. Other links point in the direction of probing the normative force of an axiom in the way exemplified by MacCrimmon and Larsson; we have dealt with this case at some length. Still other links correspond to *higher-order claims constraining the interpretation of tests*. For instance, the cognitive version of the principle that “ought” implies “can” may be invoked to blame the compound lottery axiom, rather than VNM independence, for the common consequence and common ratio effects. To handle this more disturbing category, but we suggest adding the claims to the considerations that make it possible for decision theorists to resolve the confusion of hypotheses. Here we take up Duhem’s “bon sens” while interpreting it very liberally; we boldly enlarge it to include the metaphysical and heuristic commitments that regulate the functioning of a scientific discipline. Because Duhem does not clearly say that universal Cartesian commonsense is not sufficient, and that physics would not arbitrate its internal conflicts without preconceptions of this more specific sort, our account is not exclusively moulded after him, but is only quasi-Duhemian. A whiff of Lakatosian MSRP proves to be necessary after all⁸⁴.

NOTES

* The author is grateful to Mohammed Abdellaoui for technical discussions on decision theory and to Anastasios Brenner and Brian Hill for philosophical and editorial comments.

¹ Harding's collection (1976) is the locus classicus for the so-called Duhem-Quine thesis. See also Lakatos (1970, p. 180 sq). Among those who have criticized the conflation are Vuillemin (1979), Ariew (1984), Boyer (1994), Gillies (1993, ch. 5).

² See chapters IV and V in the second part of *TP*.

³ See Fodor and Le Pore (1992) for relevant clarification. Note that the distinction between two forms of holism does not coincide with the historical difference between Duhem and Quine, because the latter, if perhaps not the former, promoted both forms.

⁴ *TP*, II, VI, §I, p. 273-278; English trans. p. 180-183.

⁵ Note the further contrast between Duhem and Quine: the latter would assuredly not restrict his underdetermination thesis to a particular group of sciences. Logically, this point is distinct from the already mentioned one that Quine extends the degree of underdetermination beyond what Duhem wishes, but the two points are entangled in Quine's exposition (e.g., *FLPV*, p. 42).

⁶ We will concentrate on falsificationism because this school has more interest in Duhem's underdetermination thesis than any other, and has made bold attempts at offering a general solution to the problem it raises. Bayesian philosophy does not seem to us to score any better, nor any alternative attempt that is currently available, but this further argument is not for this paper.

⁷ We refer to Brenner (1990) for a full discussion of how Duhem integrated the history and philosophy of science with each other.

⁸ Those of Machina (1983, 1987), Sugden (1986) and Munier (1988) still make good reading. Camerer (1995), Cohen and Tallon (2000), and Starmer (2000) extend them thoroughly, while complementing each other well.

⁹ Mongin (1997) gives an English summary of this French paper. Cross (1982) has precedence for discussing the underdetermination problem in economics, with privileged examples from macroeconomics and macroeconometrics. More papers have followed, among which Sawyer, Bid and Sankey's (1997), and they are still generally oriented towards macro-applications. Note that all this literature relies on the construal of a "Duhem-Quine thesis".

¹⁰ The title of Harding's (1974) collection, *Can Theories Be Refuted?*, is by itself indicative of the trend. But oddly enough, most of the essays skip the connection between Duhem and Popper. Among the recent textbooks in philosophy of science, Gillies's (1993) is exceptional in mentioning this connection.

¹¹ On Wiener's experiment, see *TP*, II, VI, §II, p. 279-282 (English trans., p. 184-186) and especially the items listed at the end of this passage. Foucault's experiment is discussed for the first time on p. 282-283 (English trans. p. 186-187), and Duhem returns to it several times, so it may be viewed as paradigmatic for his underdetermination thesis.

¹² See, e.g., Grünbaum's (1960) and Laudan's (1965) papers reprinted in Harding's (1976) collection.

¹³ The exception is Grünbaum (1960), who irrelevantly attributed (a) on top of (a) to Duhem. Laudan (1965) corrected him, and in doing this, proposed the now received distinction between a weak and a strong form of the "Duhem-Quine thesis" (in Harding, 1974, p. p. 159).

¹⁴ Grünbaum's (1960) recognizes that he cannot provide a criterion beyond the exclusion of the most obvious form of circularity (in Harding, 1974, p. 181).

¹⁵ Actually, despite the popular "come what may", even Quine does not seem to entertain the strong form seriously.

¹⁶ See, e.g., Lakatos (1970, p. 184). He refers to the statement of the thesis in terms of rationality as to the "strong interpretation", which is confusing given the preexisting distinction between a "weak" and "strong" form of the thesis.

¹⁷ Mongin (1988) made a similar distinction between "thèse" and "problème de Duhem".

¹⁸ For a Bayesian analysis of test and confirmation, see Howson and Urbach (1993, ch. 7).

¹⁹ This is the maximal amplification of Duhem's epistemological holism; it appears in *TP*, II, VI, §II, p. 285 and §V, p. 303-304; English trans., p. 187-188, 199-200. Brenner (1990, p. 226) suggests connecting it with Duhem's usual emphasis on the theoretical unity of physics.

²⁰ See Popper (1935-1972, § 19, n.1 and § 85, n. *2; 1963-1972, p. 112-113).

²¹ The argument is also in Duhem (1908, p. 132-133).

²² See, e.g., Laugier (1999).

²³ Both lines of reasoning can indeed be found in Popper (1963-1972, p. 238-239).

²⁴ See, e.g.: “Our new theory will represent a potential step forward, whatever the outcome of the new tests may be. For it will be better testable than the previous theory” (1963-1972, p. 242).

²⁵ See Lakatos (1970, p. 116-120). This passage defines what it means for a theory to be properly “falsified” and for a series of theories to be “progressive”. Lakatos is clearly struggling here with “the Duhem-Quine thesis” even if he does not mention it.

²⁶ Mongin (2003) discusses the sense in which this style of axiomatization meets the requirements that logicians usually put on the axiomatic method.

²⁷ See the proceedings, p. 143 and 163, as well as Samuelson (1952a). For more detail on the origin of the independence condition, see Fishburn (1989) and Fishburn and Wakker (1995).

²⁸ We skip the technical formulation. There are various definitions of continuity, the choice depending on how VNM independence is simultaneously defined. For a mathematical comparison between the systems, see Fishburn (1982).

²⁹ Compare Luce and Raiffa's approach to compound lotteries with Samuelson's (1952b, p. 671). Starting from the same definition in terms of prizes, he assumes reduction of compound lotteries to hold as a “convention”, not as substantial and possibly behavioral condition.

³⁰ It is another curious slip that the preliminary step based on (A1) and (A2) was fully clarified *after* the VNM representation theorem had been proved and become widely known.

³¹ This connection goes back to Marshall's *Principles of Economics* (1890-1921) through von Neumann and Morgenstern's *Theory of Games* (1944-1947, p. 28, 629). For a recent analysis of the pleasure of game, see Diecidue, Schmidt and Wakker (2004).

³² Later, Ellsberg (1961) came to question Savage's (1954) axioms by means of a thought-provoking experiment, and he is remembered today mostly for this contribution.

³³ For a fuller criticism, see McClennen (1983), to whom we also refer for an appraisal of the next justification.

³⁴ See in particular Hammond (1988) and Machina (1991). The first argument along Samuelson's line is to be found, even with more detail, in an unpublished paper by Rubín (1949).

³⁵ See Marshall (1890-1921, p.111-112, taken up in Mathematical Appendix, Note IX, p. 693-694).

³⁶ The hypothesis was also disputed by Markowitz (1952) on the theoretical ground that it did not properly distinguish between the individuals' current income and total wealth when a lottery is drawn out.

³⁷ Some simple applications can be found in today's microeconomics texts like Varian's (1978).

³⁸ “Fondements et applications de la théorie du risque en économétrie”, organized by Centre National de la Recherche Scientifique, 12-17 May 1952.

³⁹ Savage's *Foundations of Statistics* appeared only in 1954, but he presented some of its content in Paris, and this was hailed as an advance compared with VNM theory even by opponents to the latter, like Allais.

⁴⁰ There is a two-choice format of experiment adapted to Savage's system, and especially to the analogue in it of the VNM independence condition. This format became established in 1961 with the so-called *Ellsberg paradox*, which involves choosing between urns instead of lotteries.

⁴¹ See Jallais and Pradier's (2005) investigation of when and how the Allais paradox emerged in 1952.

⁴² The proceedings of the 1952 conference contain some applications of the two-pair format, but they are due to two other French engineers, Massé and Morlat. Their contribution anticipates on Allais's critique of EUT significantly.

⁴³ Mongin (1988) fell into the trap. Guala (2000) argued against him that normative issues had decisively influenced both Allais's conception and the others' reception of the paradox. The present account corrects the bias in the earlier one.

⁴⁴ As early as 1952, he began a vast questionnaire study, but the answers proved to be difficult to exploit, and only some results were published as late as 1979. See Allais (1979, p. 447-448) for details.

⁴⁵ See Allais (1953, p. 518). We leave out a third component, which is the use of objective probabilities, because Allais seems to vary about its rôle and meaning.

⁴⁶ At the 1952 conference, de Finetti offered it as the major argument for EUT; see the proceedings, p. 159, 196. It interacts in a complex way with the issue of the “cardinality” of the VNM function and the $u(x)$ that replaces it in Allais's positive conception. We had to leave this major topic aside from the account to keep it tractable, and instead refer to Allais's (1953a and b, 1979) discussion, along with Fishburn's (1989) and Bouyssou and Vansnick's (1990) clarifying accounts.

⁴⁷ This is fallback line. Apparently, Savage was not prepared anymore to make the strong normative claims of his 1952 paper with Friedman.

⁴⁸ For more details, see the survey part of MacCrimmon and Larsson (1979, p. 364-366). The experiment involving reconsideration was made by Slovic and Tversky (1974).

⁴⁹ Neither of these comments is made by MacCrimmon and Larsson.

⁵⁰ However, part of Allais's suppositions on the wise men remain untested. MacCrimmon and Larsson did not control for the subjects' acquaintance with the probability calculus, and in particular, for their acceptance of the compound lottery axiom.

⁵¹ Specifically, method 1, which proposes first a lottery $l = ((0.98, M), (0.02, 0))$, where M is the maximum amount of money under consideration, and then $l' = ((0.5, l), (0.5, 0)) = ((0.49, M), (0.51, 0))$ may not lead to the same result as Method 2, which starts from l' right away.

⁵² For more on the utility evaluation effect, see the surveys by Machina (1983, 1987) and Jaffray (1989).

⁵³ Hey (1991) and Camerer (1995) cover more ground. Our selection is constrained by the philosophical purpose and fixed time limits, and for expository simplicity, we left out the works, including Loomes and Sugden's (1982) and Fishburn's (1988), which relax the ordering axiom at the same time as VNM independence.

⁵⁴ See the second paper at the Paris conference (p. 127-140 of the 1953 proceedings).

⁵⁵ Allais (1953b, p. 511-513). This accords with his comments after Samuelson's and Massé and Morlat's papers (p. 154 and 194-5 of the 1953 proceedings).

⁵⁶ See Allais (1979).

⁵⁷ Hagen's formula has another possible interpretation in terms of *disappointment*, which Loomes and Sugden (1986) stress, thus connecting it with a model of their own.

⁵⁸ References are to be found in Quiggin (1982) and Camerer (1995).

⁵⁹ For the first point: the authors themselves revised "prospect theory" in the direction of RDEUT, which led to "cumulative prospect theory" (Tversky and Kahneman, 1992). For the second point: Karmakar (1978) independently put forward a S-shaped form for a to accommodate the utility evaluation effect as well as the Allais paradox, and this shape seems to have received no less experimental support than the convex shape.

⁶⁰ Technically, it is just a convex subset of a vector space, not itself a vector space.

⁶¹ Beside claiming qualitative association of risk-aversion with concavity, and of risk-love with convexity, VNM theory uses the "Arrow-Pratt index of risk-aversion" as a quantitative measure (see, e.g., Varian, 1978 for an elementary exposition). Machina's generalizations are in particular concerned with retaining the properties of this index.

⁶² Two dimensions are sufficient to represent preferences over three-outcome lotteries because of the normalization $p_1 + p_2 + p_3 = 1$. This geometric representation was introduced by Marschak (1950) to liken VNM theory with the neo-classical theories of choice under certainty.

⁶³ A stronger, but perhaps not very plausible interpretation, is that for any given pair of choices that violates VNM theory, if one assumes (H1), all local utility functions that are compatible with the violation turn out to satisfy (H2).

⁶⁴ Despite being weak in the sense explained, (H2) may be too strong in another sense, because it does not accommodate all replications of the common consequence or common ratio effects. Camerer (1995, p. 636) suggests that it should be applied to only part of the set L .

⁶⁵ Taking the particular case $u(x) = x$, Yaari (1987, p. 107) demonstrates that the f function is convex throughout the 0-1 interval if and only if the individual is risk averse in the sense of preferring any lottery l to a variant of l in which the outcomes are statistically more dispersed. For a more general theorem, see Chew, Karni and Safra (1987). Cohen and Tallon (2000) survey these results.

⁶⁶ By adopting a treatment of risk-attitudes in terms of f , they become able to use u for other semantic purposes, in particular to recover the time-honoured neo-classical hypothesis of decreasing marginal satisfaction.

⁶⁷ See the sketch in Yaari (1987, p. 104).

⁶⁸ See also Gilboa (1987) and Schmeidler (1989).

⁶⁹ See again Camerer's assessment (1995).

⁷⁰ Mongin (1988, p. 312) did not separate the two interpretations sufficiently well. Granger (1992, p. 244) suggested that the latter was more appropriate than the former at this stage of the empirical discussion, and we are following him now.

⁷¹ Compare with, e.g., the textbooks by Hey (1991) and Friedman and Sunder (1994).

⁷² See Guala (2005, ch. 7) for further discussion of the parallelism issue.

⁷³ Hausman and Mongin (1998) evidence and discuss this kind of dogmatism.

⁷⁴ Furthermore, some would add the condition that choice data be amenable to standard econometric methods. The experimental work discussed in this paper relies on straightforward counting of proportions, but the field has witnessed increased use of econometrics since the 1990s.

⁷⁵ An analogy here is that of the elector who votes for a non-preferred candidate because of the electoral system.

⁷⁶ A scientific theory T is falsified *if and only if* another theory T' has been proposed with the following characteristics: (1) T' has excess empirical content over T ...; (2) T' explains the previous success of T ...; (3) some of the excess content of T' is unrefuted" (Lakatos, 1970, p. 116, our emphasis).

⁷⁷ The textual evidence for Duhem is the already mentioned analysis of Wiener's experiment. As reported in *TP* and earlier work, it was intended to refute Neumann's hypothesis and succeeded in doing so; no alternative was involved. That Popper allows for "falsifications" without superseding theories is clear from passages such as those already discussed on "crucial arguments" in *LSD* and *Conjectures and Refutations*.

⁷⁸ Our discussion does not need Lakatos's last requirement that some of the novel predictions be "corroborated".

⁷⁹ Machina's (1983, p. 287-289) example fits this description.

⁸⁰ Mongin (1988, 1997) already argued that this demand was inapplicable to the case. The present account sharpens the critique of Popper and Lakatos, while remaining within the broad confines of refutationism.

⁸¹ See Segal (1995) and Camerer (1995) for more on this interpretation.

⁸² A fourth problematic claim, which we cannot address here, is that the new construals were developed in a normative vacuum.

⁸³ Concerning Samuelson's defence, Machina writes: "although this is a prescriptive argument, it has played a key role in economists' adoption of expected utility as a descriptive theory of choice under uncertainty" (1987, p. 127). The same could be said of Friedman and Savage's defence, despite its being flawed.

⁸⁴ Compare with the limited use of MSRP in Mongin (2002) to tackle the issue of progress in normative theories.

BIBIOGRAPHY

Fondements et applications de la théorie du risque en économétrie, Paris, Centre National de la Recherche Scientifique, 1953.

Allais Maurice (1953a), "Fondements d'une théorie positive des choix comportant un risque et critique des postulats et axiomes de l'école américaine", Appendix to *Fondements*, p. 257-332. Reprint, Paris, Imprimerie Nationale, 1955. English transl. "The Foundations of a Positive Theory of Choice Involving Risk and a Criticism of the Postulates and Axioms of the American School", in Allais and Hagen, 1979, p. 27-145.

—(1953b), "Le comportement de l'homme rationnel devant le risque", *Econometrica*, 21, p. 503-546.

—(1979), "The So-called Allais Paradox and Rational Decisions Under Uncertainty", in Allais and Hagen, 1979, p. 437-683.

—(1988), "The General Theory of Random Choices in Relation to the Invariant Cardinal Utility Function and the Specific Probability Function. The (U, θ) Model: A General Overview", in Munier, 1988, p. 231-289.

—and O. Hagen (1979) (eds), *Expected Utility Hypothesis and the Allais Paradox*, Dordrecht, D. Reidel.

Ariew Roger (1974), "The Duhem Thesis", *British Journal for the Philosophy of Science*, 35, p. 313-325.

Bouyssou Denis and Jean-Claude Vansnick (1990), "Utilité cardinale dans le certain et choix dans le risque", *Revue économique*, 6, p. 979-1000.

Boyer Alain (1994), *Introduction à la lecture de Karl Popper*, Paris, Presses de l'École normale supérieure.

Brenner Anastasios (1990), *Duhem, science, réalité et apparence*, Paris, Vrin.

Camerer Colin (1992), "Recent Tests of Generalizations of Expected Utility Theory", in Edwards, 1992, p. 207-251.

—(1995), "Individual Decision Making", in J.H. Kagel and A. Roth (eds), *The Handbook of Experimental Economics*, Princeton, Princeton University Press, ch. 8, p. 587-703.

- Chew Soo Hong, Edi Karni and Zvi Safra (1987), "Risk Aversion in the Theory of Expected Utility with Rank-Dependent Probabilities", *Journal of Economic Theory*, 42, p. 370–381.
- Cohen Michèle and Jean-Marc Tallon (2000), "Décision dans le risque et l'incertain: l'apport des modèles non-additifs", *Revue d'économie politique*, 110, p. 631-681.
- Conslík, John (1989), "Three Variants on the Allais Example", *American Economic Review*, 79, p. 392-407.
- Cross Rod (1982), "The Duhem-Quine Thesis, Lakatos and the Appraisal of Theories in Macroeconomics", *Economic Journal*, 92, p. 320-340.
- Diecidue Enrico, Ulrich Schmidt and Peter Wakker (2004), "The Utility of Gambling Reconsidered", *Journal of Risk and Uncertainty*, 29, p. 241-259.
- Duhem Pierre (1906), *La théorie physique. Son objet, sa structure*, Paris, Chevalier et Rivière (2nd ed., 1914). Reprint, Paris, Vrin, 1981, intr. P. Brouzeng. *The Aim and Structure of Physical Theory*, English transl. by P. Wiener, Princeton, Princeton University Press, 1954.
- (1908), *Sozein ta phainomena. Essai sur la notion de théorie physique de Platon à Galilée*, Paris, Hermann. Reprint Paris, Vrin, 1994, intr. by P. Brouzeng. *To Save the Phenomena*, English transl. by E. Dolland and C. Maschler, Chicago, The University of Chicago Press, 1969.
- Edwards Ward (ed.) (1992), *Utility Theories: Measurement and Applications*, Dordrecht, Kluwer.
- Ellsberg Daniel (1956), "Classic and Current Notions of 'Measurable Utility'", *Economic Journal*, 64, 528-556.
- (1961), "Risk, Ambiguity, and the Savage Axioms", *Quarterly Journal of Economics*, 75, p. 643-669.
- Fishburn Peter C. (1970), *Utility Theory for Decision Making*, New York, Wiley.
- (1982), *The Foundations of Expected Utility*, Dordrecht, Reidel, 1982.
- (1988), *Nonlinear Preference and Utility Theory*, Baltimore, Johns Hopkins University Press.
- (1989), "Retrospective on the Utility Theory of von Neumann and Morgenstern", *Journal of Risk and Uncertainty*, 2, 127–158.
- and Peter Wakker (1995), "The Invention of the Independence Condition for Preferences", *Management Science*, 41, 1130–1144.
- Fodor Jerry and Ernest Lepore (1992), *Holism: A Shopper's Guide*, Oxford, Blackwell.
- Friedman Daniel and Shyam Sunder (1994), *Experimental Methods. A Primer for Economists*, Cambridge, Cambridge University Press.
- Friedman Milton and Leonard J. Savage (1948), "The Utility Analysis of Choices Involving Risk", *Journal of Political Economy*, 56, p. 278-304.
- (1952), "The Expected Utility Hypothesis and the Measurability of Utility", *Journal of Political Economy*, 60, p. 463-474.
- Gilboa Itzhak (1987), "Expected Utility with Purely Subjective Non-Additive Probabilities", *Journal of Mathematical Economics*, 16, p. 65-88.
- Gillies Donald (1993), *Philosophy of Science in the Twentieth Century*, Oxford, Blackwell.
- Granger Gilles (1992), *La vérification*, Paris, Odile Jacob.
- Grünbaum, A. (1960), "The Duhemian Argument", *Philosophy of Science*, 27, p. 75-88, reprinted in Harding, 1974, p. 116-131.
- Guala Francesco (2000), "The Logic of Normative Falsification: Rationality and Experiments in Decision Theory", *Journal of Economic Methodology*, 7, p. 59-93.
- (2005), *The Methodology of Experimental Economics*, Cambridge, Cambridge University Press.
- Hagen, Ole (1979), "Towards a Positive Theory of Preferences under Risk", in Allais and Hagen, 1979, p. 271-302.
- Hammond Peter J. (1988), "Consequentialist Foundations for Expected Utility", *Theory and Decision*, 25, 25–78.
- Hausman Daniel and Philippe Mongin (1998), "Economists' Responses to Anomalies: Full-Cost Pricing versus Preference Reversals", in J. Davis (ed.), *New Economics and Its History. History of Political Economy, Annual Supplement*, 29, p. 255-272.
- Hey John D. (1991), *Experiments in Economics*, Oxford, Blackwell.
- Harding Sandra G. (ed.) (1976), *Can Theories be Refuted? Essays on the Duhem-Quine Thesis*, Dordrecht, Reidel.
- Jaffray Jean-Yves (1989), "Some Experimental Findings on Decision-Making under Risk and their Implications", *European Journal of Operation Research*, 38, p. 301-306.
- Jallais Sophie and Pierre-Charles Pradier (2005), "The Allais Paradox and its Immediate Consequences for Expected Utility Theory", in P. Fontaine and R. Leonard (eds), *The 'Experiment' in the History of Economics*, chap. 4, London, Routledge.

- Kahneman Daniel and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, 47, p. 263-291.
- Karmarkar Uday (1974), "The Effect of Probabilities on the Subjective Evaluation of Lotteries", WP no 698-74, MIT, Sloane School of Management.
- (1978), "Subjective Weighted Utility: A Descriptive Extension of the Expected Utility Model", *Organizational Behavior and Human Performance*, 21, p. 61-72.
- Lakatos Imre (1970), "The Methodology of Scientific Research Programmes", in I. Lakatos and A. Musgrave (eds), *Criticism and the Growth of Knowledge*, Cambridge University Press, p. 91-196.
- Laudan, L. (1965), "Grünbaum on the Duhemian Argument", *Philosophy of Science*, 32, p. 295-301, reprinted in Harding, 1974, p. 155-161.
- Laugier Sandra (1999), "Expérience cruciale", in D. Lecourt (ed.), *Dictionnaire d'histoire et de philosophie des sciences*, p. 404-406.
- Loomes Graham and Robert Sugden (1982), "Regret Theory: An Alternative Theory of Rational Choice under Uncertainty", *Economic Journal*, 92, p. 805-824.
- (1986), "Disappointment and Dynamic Consistency in Choice under Uncertainty", *Review of Economic Studies*, 53, p. 271-282.
- Luce R. Duncan and Howard Raiffa (1957), *Games and Decisions*, New York, Wiley, 1957.
- McCord Mark and Richard de Neufville (1983), "Empirical Demonstration That Expected Utility Is Not Operational", in Stigum and Wenstop, 1983, p. 181-199.
- McClennen Edward F. (1983), "Sure-Thing Doubts", in Stigum and Wenstop, 1983, p. 117-136.
- MacCrimmon Kenneth R. and Stig Larsson (1979), "Utility Theory: Axioms versus 'Paradoxes'", in Allais and Hagen, 1979, p. 333-409.
- Machina Mark J. (1982), "'Expected Utility' Analysis without the Independence Axiom", *Econometrica*, 50, p. 277-323.
- (1983), "Generalized Expected Utility Analysis and the Nature of Observed Violations of the Independence Axiom", in Stigum and Wenstop (1983, p. 263-293).
- (1987), "Choice under Uncertainty: Problems Solved and Unsolved", *Journal of Economic Perspectives*, 1, p. 121-154.
- (1991), "Dynamic Consistency and Non-Expected Utility", in M. Bacharach and S. Hurley (eds), *Foundations of Decision Theory*, p. 39-91, Oxford, Blackwell.
- Markowitz Harry (1952), "The Utility of Wealth", *Journal of Political Economy*, 60, p. 151-158.
- Marshall Alfred (1890), *Principles of Economics*, London, MacMillan, 8th ed., 1920.
- Marschak Jacob (1950), Rational Behavior, "Uncertain Prospects, and Measurable Utility", *Econometrica*, 18, p. 111-141.
- Mongin, Philippe (1988), "Problèmes de Duhem en théorie de l'utilité espérée", *Fundamenta Scientiae*, 9, p. 289-317.
- (1997), "Expected Utility Theory", in J. Davis, W. Hands and U. Mäki (eds), *Handbook of Economic Methodology*, London, Elgar, 1997, p. 342-350.
- (2003), "L'axiomatisation et les théories économiques", *Revue économique*, 54, p. 99-103.
- (2006), "A Concept of Progress for Normative Economics", *Economics and Philosophy*, 22, p. 19-54.
- Mosteller Frederick and Philip Noguee (1951), "An Experimental Measurement of Utility", *Journal of Political Economy*, 59, p. 371-404.
- Munier Bertrand (1988), "A Guide to Decision-Making under Uncertainty", in Munier, 1988, p. 1-34.
- Munier Bertrand (ed.) (1988), *Risk, Decision, and Rationality*, Dordrecht, Reidel.
- Popper Karl (1935), *Logik der Forschung*, Vienna, Julius Springer. Transl. as *The Logic of Scientific Discovery*, London, Hutchinson, 1959 (6th revised ed., 1972).
- (1963), *Conjectures and Refutations*, London, Routledge (4th revised ed., 1972).
- (1983), *The Postscript to the Logic of Scientific Discovery, I. Realism and the Aim of Science*, London, Hutchinson.
- Quiggin John (1982), "A Theory of Anticipated Utility", *Journal of Economic Behavior and Organizations*, 3, p. 323-343.
- Quine Willard V.O. (1953), *From a Logical Point of View*, Cambridge, Mass., Harvard University Press, 1980.
- Rubin, Herman (1949), "The Existence of Measurable Utility and Psychological Probability", Cowles Discussion Papers, no 332.
- Samuelson, Paul A. (1952a), "Utility, Preference, and Probability" in J. Stiglitz (ed.), *The Collected Scientific Papers of Paul A. Samuelson*, 1, ch. 13, p. 127-136.
- (1952b), "Probability, Utility, and the Independence Axiom", *Econometrica*, p. 670-678.

- Savage Leonard J. (1954), *The Foundations of Statistics*, New York, Wiley, 2nd ed., New York, Dover, 1972.
- Schmeidler David (1986), "Integral Representation without Additivity", *Proceedings of the American Mathematical Society*, 97, p. 255-261.
- (1989), "Subjective Probability and Expected Utility without Additivity", *Econometrica*, 57, p. 571-587.
- Segal Uzi (1992), "The Independence Axiom versus the Reduction Axiom: Must We Have Both?", in Edwards, 1992, p. 165-183.
- Slovic Paul and Amos Tversky (1974), "Who Accepts Savage's Axiom?", *Behavioral Science*, 19, p. 368-373.
- Starmer Chris (2000), "Developments in Non-expected Utility Theory: the Hunt for a Descriptive Theory of Choice under Risk", *Journal of Economic Literature*, 38, p. 332-82.
- Sugden Robert (1986), "New Developments in the Theory of Choice Under Uncertainty", *Bulletin of Economic Research*, 38, p. 1-24.
- Stigum Bernt and Fred Wenstop (1983) (eds), *Foundations of Utility and Risk theory with Applications*, Dordrecht, Reidel.
- Varian Hal R. (1978), *Microeconomic Analysis*, New York, Norton.
- Von Neumann John and Oskar Morgenstern (1944), *The Theory of Games and Economic Behavior*, Princeton, Princeton University Press (2nd ed. 1947).
- Vuillemin Jules (1979), "On Duhem and Quine's Theses", *Grazer Philosophische Studien*, 9, p. 69-96.
- Yaari Menahem (1987), "The Dual Theory of Choice under Risk", *Econometrica*, 55, p. 95-115.