A Simpler and More Realistic Subjective Decision Theory

Haim Gaifman* and Yang Liu†

December 1, 2016

Abstract

In his classic Savage (1954, 1972) develops a formal system of rational decision making. It is based on (i) a set of possible states of the world, (ii) a set of consequences, (iii) a set of acts, which are functions from states to consequences, and (iv) a preference relation over the acts, which represents the preferences of an idealized rational agent. The goal and the culmination of the enterprise is a representation theorem: Any preference relation that satisfies certain arguably acceptable postulates determines a (finitely additive) probability distribution over the states and a utility assignment to the consequences, such that the preferences are determined by the expected utilities of the acts. Additional problematic assumptions are however required in Savage's proofs. First, there is a Boolean algebra of events (sets of states) which determines the richness of the set of acts. The probabilities are assigned to members of this algebra. Savage's proof requires that this be a σ-algebra (i.e., closed under infinite countable unions and intersections), which makes for an extremely rich preference relation. On Savage's view we should *not* require the probability to be σ-additive. He therefore finds the insistence on a σ-algebra, peculiar and is unhappy with it. But he sees no way of avoiding it. Second, the assignment of utilities requires the *constant act assumption*: for every consequence there is a constant act, which has that consequence in every state. This is known to be highly counterintuitive. The present work includes two mathematical results. The first, and more difficult one, shows that the σ-algebra assumption can be dropped. The second states that, as long as utilities are assigned to finite gambles only, the constant act assumption can be replaced by the plausible, much weaker assumption that there are at least two non-equivalent constant acts. The paper discusses the notion of "idealized agent" that underlies Savage's approach, and argues that the simplified system, which is adequate for all the actual purposes for which the system is designed, involves a more realistic notion of an idealized agent.

Keywords. subjective probability, expected utilities, Savage axioms, realistic decision theory, partition tree, Boolean algebra.

*H. Gaifman, Columbia University, New York, USA. *Email*: hg17@columbia.edu

†Y. Liu (✉), University of Cambridge, Cambridge, UK. *Email*: yl587@cam.ac.uk

0 Introduction

Ramsey's groundbreaking work "Truth and Probability" (Ramsey, 1926) established the decision theoretic approach to subjective probability, or, in his terminology, to *degree of belief*. Ramsey's idea was to consider a person who has to choose between different practical options, where the outcome of the decision depends on unknown facts. One's decision will be determined by (i) one's probabilistic assessment of the facts, i.e., one's degrees of belief in the truth of various propositions, and (ii) one's personal benefits that are associated with the possible outcomes of the decision. Assuming that the person is a rational agent — whose decisions are determined by some assignment of degrees of belief to propositions and utility values to the outcomes — we should, in principle, be able to derive the person's degrees of belief and utilities from the person's decisions. Ramsey proposed a system for modeling the agent's point of view in which this can be done. The goal of the project is a representation theorem, which shows that the rational agent's decisions should be determined by the expected utility criterion.

The system proposed by Savage (1954, 1972) is the first decision-theoretic system that comes after Ramsey's, but it is radically different from it, and it was that system that put the decision-theoretic approach on the map.[1] To be sure, in the intervening years a considerable body of research has been produced in subjective probability, notably by de Finetti (1937a,b), and by Koopman (1940a,b, 1941), whose works are often mentioned by Savage, among many others. De Finetti also discusses problems related to expected utility. Yet these approaches were not of the decision-theoretic type: they did not aim at a unified account in which the subjective probability is derivable from decision making patterns. It might be worthwhile to devote a couple of pages to Ramsey's proposal, for its own sake and also to put Savage's work in perspective. We summarize and discuss Ramsey's work in Appendix A.

The theory as presented in Savage (1954, 1972) has been known for its comprehensiveness and its clear and elegant structure. Some researchers have considered it the best of the decision-theoretic systems. Thus Fishburn (1970) has praised it as "the most brilliant axiomatic theory of utility ever developed" and Kreps (1988) describes it as "the crowning glory of choice theory."

The system is determined by (I) The formal structure, or the basic design, and (II) The axioms the structure should satisfy, or — in Savage's terminology — the *postulates*. Savage's crucial choice of design has been to base the model on two independent coordinates: (i) a set S of *states*, (which correspond to what in other systems is the set *possible worlds*), and (ii) a set of *consequences*, $\mathbf{C}$, a new abstract construct, whose members represent the outcomes of one's *acts*. The acts themselves, whose collection is denoted here as $\mathcal{A}$, constitute the third major component. They are construed as functions from S into $\mathbf{C}$. The idea is simple: the consequence of one's act depends on the state of the world. Therefore, the act itself can be represented as a function from the set of states into the set of consequences. Thus, we can use heuristic visualization of two coordinates in a two-dimensional space.

S is provided with additional structure, namely, a Boolean algebra $\mathcal{B}$ of subsets called *events* (which, in another terminology, are *propositions*). The agent's subjective, or personal

[1] "Before this [Savage's 1954 book], the now widely-referenced theory of Frank P. Ramsey (1931) was virtually unknown." (Fishburn, 1970, p.161)

view is given by the fourth component of the system, which is a preference relation, $\succcurlyeq$, defined over the acts. All in all, the structure is:

$$(S, \mathbf{C}, \mathcal{A}, \succcurlyeq, \mathcal{B})$$

We shall refer to it as a *Savage-type decision model*, or, for short, *decision model*. Somewhat later in his book Savage introduces another important element: that of *constant acts*. It will be one of the focus points of our paper and we shall discuss it shortly. (For contrast, note that in Ramsey's system the basic component consists of propositions and worlds, where the latter can be taken as maximally consistent sets of propositions. There is no independent component of "consequences".)

Savage's notion of consequences corresponds to the "goods" in VNM — the system presented in von Neumann and Morgenstern (1944).[2] Now VNM uses gambles that are based on an objective probability distribution. Savage does not presuppose any probability but has to derive the subjective probability within his system. The most striking feature of that system is the elegant way of deriving — from his first six postulates — a (finitely additive) probability over the Boolean algebra of events. That probability is later used in defining the utility function, which assigns utilities to the consequences. The definition proceeds along the lines of VNM, but since the probability need not be σ-additive, Savage cannot apply directly the VNM construction. He has to add a seventh postulate and the derivation is somewhat involved.

In this paper we assume some familiarity with the Savage system. For the sake of completeness we include a list of the postulates in Appendix B.

As far as the postulates are concerned Savage's system constitutes a very successful decision theory, including a decision-based theory of subjective probability. Additional assumptions, which are not stated as axioms, are however required both (i) in Savage's derivation of subjective probability and (ii) in his derivation of subjective utility. These assumptions are quite problematic and our goal here is to show how they can be eliminated, and how the elimination yields a simpler and more realistic theory.

The first problematic assumption we address in this paper is the *σ-algebra assumption*: In deriving the subjective probability, Savage has to assume that the Boolean algebra, $\mathcal{B}$, over which the probability is to be defined is a σ-algebra (i.e., closed under countable infinite unions and intersections). Now Savage insists that we should *not* require that the probability be σ-additive. He fully recognizes the importance of the mathematical theory of probability, which is based on Kolmogorov's axioms, according to which $\mathcal{B}$ is a σ-algebra and the probability is σ-additive. But he regards it as a *mathematical* theory, and σ-additivity — as a sophisticated mathematical concept. Being a rational agent should not require the capacities of a professional mathematician. In this, Savage follows de Finetti (both made important mathematical contributions to mathematical probability theory). It is therefore odd that the Boolean algebra, over which the finitely additive probability is to be defined, is required to be a σ-algebra. Savage notes this oddity and justifies it on grounds of expediency, he sees no other way of deriving the quantitative probability that is needed for the purpose of defining expected utilities

[2] At the time Savage served as chief "statistical" assistant to von Neumann.

> It may seem peculiar to insist on σ-algebra as opposed to finitely additive algebras even in a context where finitely additive measures are the central object, but countable unions do seem to be essential to some of the theorems of §3 — for example, the terminal conclusions of Theorem 3.2 and Part 5 of Theorem 3.3. (Savage, 1972, p.43)

The theorems he refers to are exactly the two places (see Savage, 1972, p.37-38) where his proof relies on the σ-algebra assumption. Assuming the σ-algebra assumption he shows the existence of a unique finitely additive numeric probability in systems that satisfy his axioms. He also shows that this probability satisfies a certain property, let us call it "completeness".[3] Completeness guarantees that the probability can serve as a basis for assigning utilities to the consequences, so that the expected utilities play their usual role. Using again the σ-algebra assumption he shows that the probability he defined is complete.

We eliminate the σ-algebra assumption first by pointing a way of defining probabilities, in systems satisfying Savage's axioms, which does not rely on the σ-algebra assumption. This is the hard technical core of this paper and about a third of it. We develop, for the purpose of the proof a new technique based on what we call *tri-partition trees*. This is done in Section 2.

We also show that there is a condition weaker than completeness, call it "almost completeness" that is sufficient for assigning utilities to consequences, and that the probabilities we construct are almost complete. We show also that, without the σ-algebra assumption, there is no way of getting complete probabilities, but under the σ-algebra assumption every almost complete probability is complete. Thus Savage was right when he thought that the the σ-algebra assumption is necessary for "the terminal conclusions of Theorem 3.2". He was also right that without the σ-algebra assumption there is no way of getting the result of part 5 of Theorem 3.3. But Part 5 is not the final part, and Savage was wrong in thinking that the terminal conclusion of Part 7 requires the σ-algebra assumption.

The second problematic assumption concerns constant acts, where an act f is constant if for some fixed consequence $a \in \mathbf{C}$, $f(x) = a$, for all $x \in S$. Let $\mathfrak{c}_a$ denote that act. Note that, in Savage's framework, the utility-value of a consequence depends only on the consequence, not on the state in which it is obtained. Hence, the preorder of the constant acts, induces a preorder of the corresponding consequences:

$$a \geq b \quad \iff_{\text{Df}} \quad \mathfrak{c}_a \succcurlyeq \mathfrak{c}_b$$

where a, b range over all consequences for which there exist a corresponding constant act for which $\mathfrak{c}_x$ exists. The *Constant Acts Assumption* (CAA) is:

CAA: For every consequence $a \in \mathbf{C}$ there exists a constant act $\mathfrak{c}_a \in \mathcal{A}$.

Savage does not state CAA explicitly, but it is clearly implied by his discussion and it is needed in his proof of the representation theorem. Note that if CAA holds then the above induced preorder is a total preorder of $\mathbf{C}$.

[3] This is not Savage's terminology; he does not give this property a name.

In what follows a *simple act* is an act with a finite range of values. The term used by Savage (1972, p.70) is 'gamble'; he defines it as an act, f, such that, for some finite set, A, $f^{-1}(A)$ has probability 1. It is easily seen that an act is a gamble iff it is equivalent to a simple act. 'Gamble' is also used in gambling situations, where one accepts or rejects bets. We shall use 'simple act' and 'gamble' interchangeably. Using the probability that has been obtained already, the following is derivable from the first six postulates and CAA.

Proposition 0.1 (Simple Act Utility)**.** We can associate utilities with all consequences, so that, for all simple acts the preference is determined by the acts' expected utilities.[4]

CAA has however highly counterintuitive implications, a fact that has been observed by several scholars.[5] The consequences of a person's act depend, as a rule, on the state of the world. More often than not, a possible consequence in one state is impossible in another. Assume that I have to travel to a nearby city and can do this either by plane or by train. At the last moment I opt for the plane, but when I arrive at the airport I find that the flight has been canceled. If a and b are respectively the states *flight-as-usual* and *flight-canceled*, then the consequence of my act in state a is something like 'arrived at X by plane at time Y'. This consequence is impossible — logically impossible, given the laws of physics — in state b. Yet CAA implies that this consequence, or something with the same utility-value, can be transferred to the state b.[6] Our result concerning CAA shows that it can be avoided at some price, which — we later shall argue — is worth paying. To state the result, let us first define *feasible consequences*: A consequence a is *feasible* if there exists some act, $f \in \mathcal{A}$, such that $f^{-1}(a)$ is not a null event.[7] It is not difficult to see that the name is justified and that unfeasible consequences, while theoretically possible, are a pathological curiosity. Note that if we assume CAA then all consequences are trivially feasible. Let us replace CAA by the following much weaker assumption:

2CA: There are two non-equivalent constant acts $\mathfrak{c}_a$ and $\mathfrak{c}_b$.

(Note that 2CA makes the same claim as postulate P5; but this is misleading. While P5 presupposes CAA, 2CA does not.) Having replaced CAA by 2CA we can prove the following:

Proposition 0.2 (Simple Act Utility*)**.** We can associate utilities with all feasible consequences, so that, for all simple acts, the preference is determined by the acts' expected utilities.

[4]In order to extend that proposition to all acts, Savage adds his last postulate, P7. See also Fishburn (1970, Chapter 14) for a detailed presentation.

[5]Fishburn (1970) who observe that CAA is required for the proof of the representation theorem, has also pointed out its problematic nature. This difficulty was also noted by Luce and Krantz (1971), Pratt (1974), Seidenfeld and Schervish (1983), Shafer (1986), Joyce (1999), among others.

[6] Fishburn (1970, p.166-7), went into the problem at some detail. He noted that, if $W(x)$ is the set of consequences that are possible in state x, then we can have $W(s) \neq W(s')$, and even $W(s) \cap W(s') = \varnothing$. He noted that, so far there is no proof that avoids CAA, and suggested a line of research that would enrich the set of states by an additional structure, (see also Fishburn, 1981, p.162). The decision model in Gaifman and Liu (2015) (also sketched in Section 3) avoids the need for an additional structure, as far as simple acts are concerned.

[7]A null event is an event B, such that, given B, all acts are equivalent. These are the events whose probability is 0. See also Appendix B.

It is perhaps possible to extend this result to all acts whose consequences are feasible. This will require a modified form of P7 (there is a natural candidate for the role, which we have not tried yet). But our proposed modification of the system does not depend on there being such an extension. In our view the goal of a subjective decision theory is to handle all scenarios of having to choose from a finite number of options, involving altogether a finite number of consequences. (Simple Act Utility*) is therefore sufficient. The question of extending it to all feasible acts is intriguing because of its mathematical interest, but this is a different matter.

The rest of the paper is organized as follows. In what immediately follows we introduce some further concepts and notations which will be used throughout the paper. Section 1 is devoted to the analysis of *idealized rational agents* and what being "more realistic" abut it entails. We argue that, when carried too far, the idealization voids the very idea underlying the concept of *personal* probability and utility; the framework then becomes, in the best case, a piece of abstract mathematics. Section 2 is devoted to the σ-algebra assumption. It consists of a short overview of Savage's original proof followed by a presentation of the tri-partition trees and our proof, which is most of the section. In Section 2.3, we outline a construction by which, from a given finite decision model that satisfies P1-P5, we get a countable infinite decision model that satisfied P1-P6; this model is obtained as a direct limit of an ascending sequence of finite models. In Section 3, we take up the problem of CAA. We argue that, as far as decision theory is concerned, we need to assign utilities only to simple acts. Then we indicate the proof of Proposition 0.2. To a large extent this material has been presented in Gaifman and Liu (2015), hence we contend ourselves with a short sketch.

Some Terminologies, Notations, and Constructions

Recall that '$\succcurlyeq$' is used for the preference relation over the acts. $f \succcurlyeq g$ says that f is equi-or-more preferable to g; $\preccurlyeq$ is its converse. $\succcurlyeq$ is a preorder, which means that it is a reflexive and transitive relation; it is also *total*, which means that for every f, g either $f \succcurlyeq g$ or $g \succcurlyeq f$. If $f \succcurlyeq g$ and $f \preccurlyeq g$ then the acts are said to be equivalent, and this is denoted as $f \equiv g$. The strict preference is defined by: $f \succ g \iff f \succcurlyeq g$ and $g \not\succcurlyeq f$; its converse is $\prec$.

For a given consequence a, $\mathfrak{c}_a$ is the constant act whose consequence is a for all states. This notation is employed under the assumption that such an act exists. If $\mathfrak{c}_a \succcurlyeq \mathfrak{c}_b$ then we put: $a \geq b$. Similarly for strict preference. Various symbols are used under systematic ambiguity, e.g., '$\equiv$' for acts and for consequences, '$\leq$', '$<$' for consequences as well as for numbers. Later, when qualitative probabilities are introduced, we shall use $\succeq$ and $\preceq$, for the "greater-of-equal" relation and its converse, and $\succ$ and $\prec$ for the strict inequalities.

Note. Following Savage, by a *numeric probability*, we mean in this paper a finitely additive probability function. If σ-additivity is meant this will be explicitly indicated.

Cut-and-Paste: If f and g are acts and E is an event then we define

$$(f|E + g|\overline{E})(s) =_{\text{Df}} \begin{cases} f(s) & \text{if } s \in E \\ g(s) & \text{if } s \in \overline{E}, \end{cases}$$

where $\overline{E} = S - E$ is the complement of E.[8].

Note that $f|E + g|\overline{E}$ is obtained by "cutting and pasting" parts of f and g, which results in the function that agrees with f on E, and with g on $\overline{E}$. Savage takes it for granted that the acts are closed under cut-and-paste. Although the stipulation is never stated explicitly, it is obviously a property of $\mathcal{A}$. It is easily seen that by iterating the cut-and-paste operations just defined we get a cut-and-paste that involves any finite number of acts. It is of the form:

$$f_1|E_1 + f_2|E_2 + \ldots + f_n|E_n, \text{ where } \{E_1, \ldots, E_n\} \text{ is a partition of } S.$$

1 The Logic of the System and the Role of "Idealized Rational Agents"

The decision theoretic approach construes a person's subjective probability in terms of its function in determining the person's decision under uncertainty. The uncertainty should however stem from lack of empirical knowledge, not from one's limited deductive capacities. One could be uncertain because one fails to realize that such and such facts are logically deducible from other known facts. This type of uncertainty does not concern us in the context of subjective probability. Savage (1972, p.7) therefore posits an idealized person, with unlimited deductive powers in logic, and he notes (in a footnote on that page) that such a person should know the answers to all decidable mathematical propositions. By the same token, we should endow our idealized person with unlimited computational powers. This is of course unrealistic; if we *do* take into account the rational agent's bounded deductive, or computational resources, we get a "more realistic" system. This is the sense that Hacking (1967) meant in his "A slightly more realistic personal probability"; a more recent work on that subject is Gaifman (2004). But this is *not* the sense of "realistic" of the present paper. By "realistic", we mean *conceptually* realistic; that is, the ability to conceive impossible fantasies and treat them as if they were real.

We indicated in the introduction that CAA may give rise to agents that have such extraordinary powers of conceiving. We shall elaborate on this sort of unrealistic abilities shortly. The σ-algebra assumption can lead to even more extreme cases in a different area: the foundation of set theory. We will not go into this here, since this would require too long a detour, but we shall discuss it briefly in Section 2.1.

It goes without saying that the extreme conceptual unrealism, of the kind we are considering here, has to be distinguished from the use of hypothetical mundane scenarios — the bread-and-butter of every decision theory that contains more than experimental results. Most, if not all, the scenarios treated in papers and books of decision theory are hypothetical, but sufficiently grounded in reality. The few examples Savage discusses in his book are of this kind. The trouble is that the solutions that he proposes require that the agent be able to assess the utilities of physical impossibilities and to weigh them on a par with everyday opportunities.

Let us consider a simple decision problem, an illustrative example proposed by Savage (1972, p.13-14), which will serve us for more than one purpose. We shall refer to it as *Omelet*.

[8] Some writers use '$f \oplus_E g$' or '(f, E, g)' or 'fEg' or '$[f \text{ on } E, \text{g on } \overline{E}]$'.

John (in Savage (1972) he is "you") has to finish making an omelet began by his wife, who has already broken into a bowl five good eggs. A sixth unbroken egg is lying on the table, and it must be either used in making the omelet, or discarded. There are two states of the world *good* (the sixth egg is good) and *rotten* (the sixth egg is rotten). John considers three possible acts, f_1: break the sixth egg into the bowl, f_2: discard the sixth egg, f_3: break the sixth egg into a saucer; add it to the five eggs if it is good, discard it if it is rotten. The consequences of the acts are as follows:

$f_1(good) =$	six-egg omelet	$f_1(rotten) =$	no omelet and five good eggs wasted
$f_2(good) =$	five-egg omelet and one good egg wasted	$f_2(rotten) =$	five-egg omelet
$f_3(good) =$	six-egg omelet and a saucer to wash	$f_3(rotten) =$	five-egg omelet and a saucer to wash

Omelet is one of the many scenarios in which CAA is highly problematic. It requires the existence of an act by which a good six-egg omelet is made out of five good eggs and a rotten one.[9] Quite plausibly, John can imagine a miracle by which a six-egg omelet is produced from five good eggs and a rotten one; this lies within his conceptual capacity. But this would not be sufficient; he has to take the miracle seriously enough, so that he can rank it on a par with the other real possibilities, and eventually assign to it a utility value. This is what the transfer of *six-egg omelet* from state *good* to state *rotten* means. In another illustrative example (Savage, 1972, p.25) the result of such a transfer is that the person can enjoy a refreshing swim with her friends, while in fact she is "...sitting on a shadeless beach twiddling a brand-new tennis racket" — because she bought a tennis racket instead of a bathing suit — "while her [one's] friends swim." CAA puts extremely high demands on what the agent, even an idealized one, should be able to conceive.

CAA is the price Savage has to pay for making the consequences completely independent of the states.[10] A concrete consequence is being abstracted so that only its personal value remains. These values can be then smoothly transferred from one state to another. Our suggestion for avoiding it is described in the introduction. In Section 3 we shall argue that the price one has to pay is worth paying.

Returning to *Omelet*, let us consider how John will decide. It would be wrong to describe him as appealing to some intuitions about his preference relation, or interrogating himself about it. John determines his preferences by appealing to his intuitions about the likeliness of the states and the personal benefits he might derive from the consequences.[11] If he thinks that *good* is very likely and washing the saucer — in the case of *rotten* — is rather bothersome,

[9] "Omelet" obviously means a good omelet.

[10] This price is avoided in (Ramsey, 1926) because for Ramsey the values derive from the propositions and, in the final account, from the states. CAA is also avoided in Jeffrey (1965, 1983), because the Jeffrey-Bolker system realizes, in a better and more systematic way, Ramsey's point of view. That system however is of a different kind altogether, and has serious problems of its own, which we shall not address here.

[11] The preference relation is not "given" in the same way that the entrenched notion of probability, with its long history, is. The preference relation is rather a tool for construing probability in a decision theoretic way. John can clarify to himself what he means by "more probable" by considering its implications for making

he will prefer f_1 to the other acts; if washing the saucer is not much bother he might prefer f_3; if wasting a good egg is no big deal, he might opt for f_2.

If our interpretation is right, then a person derives his or her preferences by combining subjective probabilities and utilities. On the other hand, the representation theorem goes in the opposite direction: from preference to probability and utility. As a formal structure, the preference relation is, in an obvious sense, more elementary than a real valued function. If it can be justified directly on rationality grounds, this will yield a normative justification to the use probability and utility.

The Boolean algebra in *Omelet* is extremely simple; besides S and $\varnothing$ it consists of two atoms. The preference relation implies certain constraints on the probabilities and utility values, but it does not determine them. This, as a rule, is the case whenever the Boolean algebra is finite.[12] Now the idea underlying the system is that if the preference relation is defined over a sufficiently rich set of acts (and if it satisfies certain plausible postulates) then both probabilities and utilities are derivable from it. As far as probability is concerned, the consequences play a minor role. We need only two non-equivalent constant acts, say $\mathfrak{c}_a, \mathfrak{c}_b$, and we need only the preferences over two-valued act, in which the values are a or b. But $\mathcal{B}$ has to satisfy P6′, which implies that is must be infinite, and — in Savage's proof — it should be a σ-algebra, which implies that its cardinality is $2^{\aleph_0}$ at least, and it can be quite a complicated structure. Our result makes it possible to get a countable Boolean algebra, $\mathcal{B}$, and a decision model $(S, \mathbf{C}, \mathcal{A}, \succcurlyeq, \mathcal{B})$ which is a direct limit of an ascending sequence of substructures $(S_i, \mathbf{C}, \mathcal{A}_i, \succcurlyeq_i, \mathcal{B}_i)$, where the S_i's are finite, and where $\mathbf{C}$ is any fixed set of consequences containing two non-equivalent ones. This construction is described briefly at the end of the next section.

2 Eliminating the σ-Algebra Assumption

2.1 Savage's Derivation of Numeric Probabilities

Savage's derivation of a numeric probability comprises two stages. First, he defines, using P1-P4 and the assumption that there are two non-equivalent constant acts, a *qualitative probability*. This is a binary relation, $\succeq$, defined over events, which satisfies the axioms proposed by de Finetti (1937a) for the notion "X is more-than-or-equi probable than Y". The second stage is devoted to showing that if a qualitative probability, $\succeq$, satisfies certain additional assumptions, then there is a unique numeric probability, μ, that represents $\succeq$, that is, for all events E, F:

$$E \succeq F \iff \mu(E) \geq \mu(F) \tag{2.1}$$

practical decisions. In a more operational mood one might accord the preference relation a self-standing status. Whether Savage is inclined to this is not clear. He *does* appeal to intuitions about the probabilities; for example, in comparing P6′ to an axiom suggested by de Finetti and by Koopman, he argues that it is more intuitive, (and we agree with him). This is even clearer with regard to P6 — the decision-theoretic analog of P6′ which implies P6′.

[12] This is the case even if the number of consequence is infinite. There are some exceptions: if $\mathfrak{c}_a and \mathfrak{c}_b$ are non-equivalent and if E and E' are two events then the equivalence: $\mathfrak{c}_a|E + \mathfrak{c}_b|E' \equiv \mathfrak{c}_b|E' + \mathfrak{c}_a|E$ implies that E and E' have equal probabilities. Using equivalences of this form makes it possible to determine certain probability distributions over a finite set of atoms.

Our improvement on Savage's result concerns only the second stage. For the sake of completeness we include a short description of the first.

2.1.1 From Preferences over Acts to Qualitative Probabilities

The qualitative probability, $\succeq$, is defined by:

Definition 2.1. For any events E, F, say that E is *weakly more probable* than F, written $E \succeq F$ (or $F \preceq E$), if, for any $\mathfrak{c}_a$ and $\mathfrak{c}_b$ satisfying $\mathfrak{c}_a \succ \mathfrak{c}_b$, we have

$$\mathfrak{c}_a|E + \mathfrak{c}_b|\overline{E} \succcurlyeq \mathfrak{c}_a|F + \mathfrak{c}_b|\overline{F}. \tag{2.2}$$

E and F are said to be *equally probable*, in symbols $E \equiv F$, if both $E \succeq F$ and $F \succeq E$.

Savage's P4 guarantees that the above concept is well defined, i.e., (2.2) does not depend on the choice of the pair of constant acts. The definition has a clear intuitive motivation and it is not difficult to show that $\succeq$ is a qualitative probability, as defined by de Finetti:

Definition 2.2 (Qualitative probability)**.** A binary relation $\succeq$ over $\mathcal{B}$ is said to be a *qualitative probability* if the following hold for all $A, B, C \in \mathcal{B}$:

i. $\succeq$ is a total preorder,

ii. $A \succeq \varnothing$,

iii. $S \succ \varnothing$,

iv. if $A \cap C = B \cap C = \varnothing$ then

$$A \succ B \iff A \cup C \succ B \cup C. \tag{2.3}$$

where $\succ$ is the strict (i.e., the asymmetric) part of $\succeq$.

For a given a decision model, which satisfies P1-P4 and which has two non-equivalent constant acts, the *qualitative probability of the model* is the qualitative probability defined via Definition 2.1. If that qualitative probability is representable by a quantitative probability, and if moreover the representing probability is unique, then we get a single numeric probability and we are done.[13] The following postulate ascribes to the qualitative probability the property which Savage (1972, p.38) suggests as the key for deriving numeric probabilities.

P6′: For all events E, F, if $E \succ F$, then there is a partition $\{P_i\}_{i=1}^n$ of S such that $E \succ F \cup P_i$ for all $i = 1, \ldots, n$.

[13] Some such line of thought has guided de Finetti (1937a). Counterexamples were however found of qualitative probabilities that are not representable by any quantitative one. First to be found were counterexamples in which the Boolean algebra is infinite. They were followed by counterexamples for the finite case, in particular, a counterexample in which the qualitative probability is defined over the Boolean algebra of all subsets of a set consisting of 5 members, (cf. Kraft et al., 1959).

P6$'$ is not stated in terms of $\succcurlyeq$, which is a preference relation over acts. But, given the way in which the qualitative probability has been defined in terms of $\succcurlyeq$, P6$'$ is obviously implied by P6 (see Appendix B). As Savage describes it, the motivation for P6 is its intuitive plausibility and its obvious relation to P6$'$.

Before proceeding to the technical details that occupy most of this section it would be useful to state for comparison the two theorems, Savage's and ours, and pause on some details regarding the later use of the probabilities in the derivation of utilities.

Overview of the Main Results

We state the results as theorems about qualitative probabilities. The corresponding theorems within the Savage framework are obtained by replacing the qualitative probability $\succeq$ by the preference relation over acts $\succcurlyeq$, and P6$'$ — by P1-P6.

Theorem 2.3 (Savage)**.** Let $\succeq$ be a qualitative probability defined over the Boolean algebra $\mathcal{B}$. If (i) $\succeq$ satisfies P6$'$ and (ii) $\mathcal{B}$ is a σ-algebra, then there is a unique numeric probability μ, defined over $\mathcal{B}$, which represents $\succeq$. That probability has the following property:

(†) For every event, A, and every $\rho \in (0,1)$, there exists an event $B \subseteq A$ such that $\mu(B) = \rho \cdot \mu(A)$.

Theorem 2.4 (Main Theorem)**.** Let $\succeq$ be a qualitative probability defined over the Boolean algebra $\mathcal{B}$. If $\succeq$ satisfies P6$'$, then there is a unique numeric probability μ, defined over $\mathcal{B}$, which represents $\succeq$. That probability has the following property:

(‡) For every event, A, every $\rho \in (0,1)$, and every $\epsilon > 0$ there exists an event $B \subseteq A$, such that $(\rho - \epsilon) \cdot \mu(A)) \leq \mu(B) \leq \rho \cdot \mu(A)$.

Remark 2.5. (1) In the Introduction we used "complete" for characterizing Boolean algebras satisfying (†) and "almost complete" for those satisfying (‡).

(2) Given a numeric probability μ, let a *ρ-portion* of an event A be any event $B \subseteq A$ such that $\mu(B) = \rho \cdot \mu(A)$. Then (†) means that, for every $0 < \rho < 1$, every event has a ρ-portion. (‡) is a weaker condition: for every A, and for every $\rho \in (0,1)$, there are ρ'-portions of A, where ρ' can be $< \rho$ and arbitrary near to it.

(†) — for the case $A = S$ — implies that the set of values of μ is the full interval $[0,1]$. But (‡) only implies that the set of values is dense in $[0,1]$. Obviously, the satisfaction of P6$'$ implies that the Boolean algebra is infinite, but, as indicated in Section 2.3 it can be countable, in which case (†) must fail.

(†) is stated as one of the claims of Theorem 2 in Savage (1972, p.34); it is later used when, on the basis of the probability, utilities are assigned to the consequences, so as to yield eventually the expected utility theorem. The proof of (†) relies crucially on the σ-algebra assumption. Indeed, we have shown (by constructing a counterexample) that without the σ-algebra assumption (†) can fail. Yet, as we will show in Section 3, we can assign utilities to the consequences if we have (‡). The proof of (‡) is given in Section 2.2.4 and, as we shall see, it does not rely on the σ-algebra assumption.

2.1.2 Savage's Proof

The proof is given in the more technical part of the book (Savage, 1972, p.34-38). The presentation seems to be based on working notes, reflecting a development that led Savage to P6′. Many proofs consists of numbered claims and sub-claims, whose proofs are left to the reader (some of these exercises are difficult). Some of the theorems are supposed to provide motivation for P6′, which is introduced (on p.38) after the technical part: "In the light of Theorems 3 and 4, I tentatively propose the following postulate ...". Some of the concepts that Savage employs have only historical interest. While many of these concepts are dispensable if P6′ is presupposed, some remain useful for clarifying the picture and are therefore used in later textbooks, (e.g., Kreps, 1988, p.123). We shall use them as well.

Definition 2.6 (fine)**.** A qualitative probability is *fine* if for every $E \succ \varnothing$ there is a partition $S = P_1 \cup ... \cup P_n$ such that $E \succ P_i$, for every $i = 1, ..., n$.

Another useful concept is *tight*:

Definition 2.7 (tight)**.** A qualitative probability is *tight*, if whenever $E \succ F$, there exists $C \succ \varnothing$, such that $E \succ F \cup C \succ F$.

Obviously (fine) is the special case of P6′, where the smaller set is $\varnothing$. It is easy to show that P6′ $\iff$ (fine) + (tight). In this "decomposition" (tight) is "exactly" what is needed in order to pass from (fine) to P6′.

Remark 2.8. Savage's definition of "tight" (p.34) is different from our *tight*, it is more complicated and has only historical interest, although the two are equivalent if we presuppose (fine). Let us say that the probability function μ *almost represents* $\succeq$ (in Savage's terminology "almost agrees with" $\succeq$) if, for all E, F

$$E \succeq F \Longrightarrow \mu(E) \geq \mu(F) \tag{2.4}$$

Since $E \not\succeq F \Rightarrow F \succ E$ it is easily seen that if μ almost represents $\succeq$ then it represents $\succeq$ iff

$$E \succ F \Longrightarrow \mu(E) > \mu(F) \tag{2.5}$$

Savage's proof presupposes (fine), and its upshot is the existence of a unique μ that almost represents $\succeq$. Now (fine) implies that if $E \succ \varnothing$, then $\mu(E) > 0$. (Let $P_1, ..., P_n$ be a partition of S such that $P_i \preceq E$ for all $i = 1, ..., n$. For some i, $\mu(P_i) > 0$, otherwise $\mu(S) = 0$. Hence $\mu(E) > 0$). With (tight) added, this implies (2.5). Hence, under P6′, μ is the unique probability representing $\succeq$.

Savage's proof can be organized into three parts. Part I introduces the concept of an *almost uniform partition*, which plays a central role in the whole proof, and proves the theorem that links it to the existence of numeric probabilities. Before proceeding recall the following:

Definition 2.9. (i) A *partition* of B is a collection of disjoint subsets of B, referred to as *parts*, whose union is B. We presuppose that the number of parts is > 1 and is finite and that B is non-null, i.e., $B \succ \varnothing$.

(ii) It is assumed that no part is a null-event, unless this is explicitly allowed.

(iii) By an *n-partition* we mean a partition into n parts (this is what Savage calls *n-fold* partition).

(iv) We adopt self-explanatory expression, such as, "a partition $A = A_1 \cup ... \cup A_n$", which means that the sets on the right-hand side are a partition of A.

Definition 2.10. An *almost uniform partition* of an event B is a partition of B into a finite number of disjoint events, such that the union of any $r+1$ parts has greater or equal qualitative probability than the union of any r parts. An *almost uniform n-partition* of B is an n-partition of B which is almost uniform.

The main result of Part I comprises what in Savage's enumeration are Theorem 1 and its proof, and the first claim of Theorem 2 (on the bottom of p.34), and *its* proof. The latter consists of steps 1-7 and ends in the middle of p.36. All in all, the result in Part I is:

Theorem 2.11. If for infinitely many ns there are almost uniform n-partitions of S, then there exists a unique numerical probability, μ, which almost represents $\succeq$.

The proof of this result consists mainly of direct computational/combinatorial arguments; it is given with sufficient details and does *not* use the σ-algebra assumption. We shall take the theorem and its proof for granted.

Part II consists in showing that (fine) and the σ-algebra assumption imply that there are almost uniform n-partitions for infinitely many ns (together with the theorems of part I this yields a unique probability that almost represents the qualitative one). This is done in Theorem 3. The theorem consists of a sequence of claims, referred to as "parts", in which later parts are to be derived from earlier ones. The arrangement is intended to help the reader to find the proofs. For the more difficult parts additional details are provided. Many claims are couched in terms that have only historical interest. For our purposes, we need only to focus on a crucial construction that uses what we shall call "iterated 3-partitions" and to which we shall shortly return. This construction is described in the proof of Part 5 (on the top of p.35). As a last step it involves the crucial use of the σ-algebra assumption.

Part III of Savage's proof consists in the second claim of the aforementioned Theorem 2. It asserts that the numeric probability, which is derivable from the existence of almost uniform n-partitions for arbitrary large ns, satisfies (†). The proof consists in three claims, 8a, 8b, 8c, the last of which relies on on the σ-algebra assumption. The parallel part of our proof is the derivation of (‡), without using that assumption.

2.1.3 Savage's Method of Iterated 3-Partitions

In order to prove Part 5 of Theorem 3, Savage claims that the following is derivable from the laws of qualitative probabilities and (fine).

Theorem 2.12 (Savage)**.** For any given $B \succ \varnothing$ there exists an infinite sequence of 3-partitions of B: $\{C_n, D_n, G_n\}_n$, which has the following properties:[14]

(1) $C_n \cup G_n \succeq D_n$ and $D_n \cup G_n \succeq C_n$

(2) $C_n \subseteq C_{n+1}$, $D_n \subseteq D_{n+1}$, hence $G_n \supseteq G_{n+1}$

(3) $G_n - G_{n+1} \succeq G_{n+1}$

These properties imply that G_n becomes arbitrary small as $n \to \infty$, that is:

(4) For any $F \succ \varnothing$, there exists n such that $G_m \prec F$ for all $m \geq n$.

Note. Condition (3) in Theorem 2.12 means that G_n is a disjoint union of two subsets, $G_n = G_{n+1} \cup (G_n - G_{n+1})$, each of which is $\succeq G_{n+1}$. In this sense G_{n+1} is less than or equal to "half of G_n". Had the probability been numeric we could have omitted the scare quotes; it would have implied that the probabilities of G_n tend to 0, as $n \to \infty$. In the case of a qualitative probability the analogous conclusion is that the sets become arbitrary small, in the non-numerical sense.

Savage provides an argument, based on (fine), which derives (4) from the previous properties. The argument is short and is worth repeating: Given any $F \succ \varnothing$, we have to show that, for some n, $G_n \prec F$. Assume, for contradiction, that this is not the case. Then $F \preceq G_n$, for all ns. Now (fine) implies that there is a partition $S = P_1 \cup ... \cup P_m$ such that $P_i \preceq F$, for $i = 1, ..., m$. If $F \preceq G_n$, then $P_1 \preceq G_n$, hence $P_1 \cup P_2 \preceq G_{n-1}$, hence $P_1 \cup P_2 \cup P_3 \cup P_4 \preceq G_{n-2}$, and so on. Therefore, if $2^{k-1} \geq m$, then $S \preceq G_1$, which is a contradiction.

Definition 2.13. Call an infinite sequence of 3-partitions of B, which satisfies conditions (1), (2), (3), a *Savage chain for B*. We say that the chain *passes through* a 3-partition of B, if the 3-partition occurs in the sequence.

We presented the theorem so as to conform with Savage's notation and the capital letters he used. Later we shall change the notation. We shall use ordered triples for the 3-partition and place in the middle the sets that play the role of the G_ns. The definition just given can be rephrased of course in terms of our later terminology.

Figure 2.1 is an illustration of the construction of a Savage chain. Presenting the Savage chain as a sequence of triples with the G_ns in the middle, makes for better pictorial representation. And it is essential when it comes to trees.

The fact that $D_n \cup G_n \succeq C_n$, $C_n \cup G_n \succeq D_n$, and the fact that G_n becomes arbitrary small suggest that G_n plays the role of a "margin of error" in a division of the set into two,

[14] The proof of the existence of such a sequence is left to the reader. Fishburn (1970, pp.194-197) reconstructs parts of Savage's work, filling in missing segments. Part 5 of Theorem 3 is among the material Fishburn covers. Fishburn presupposes however a qualitative probability that satisfies P6′ (F5 – in his notation). Therefore his proof cannot be the one meant by Savage; the latter uses only (fine). We believe that it should not be too difficult to make such a proof, or to modify Fishburn's proof of part 5, so as to get a proof from (fine) only. The matter is not too important, since the problem of the σ-algebra assumption concerns qualitative logic that satisfies P6′. Besides, we can trust Savage that his claims are derivable from (fine) alone.

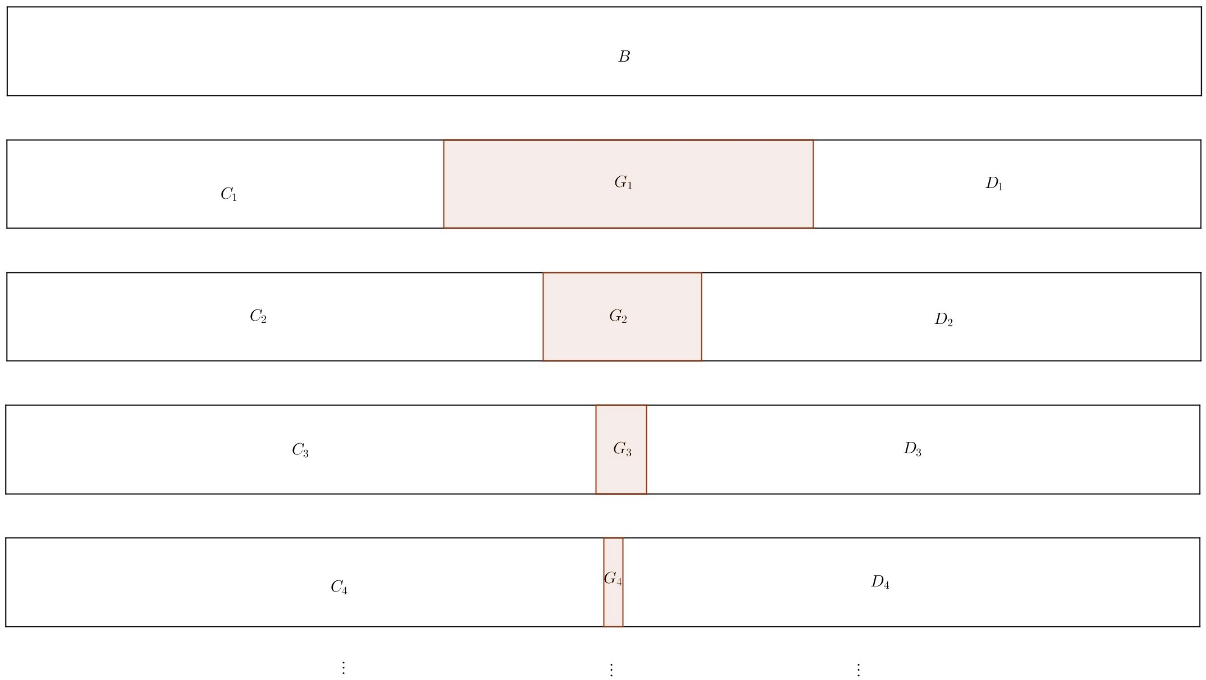

Figure 2.1: Savage's error reducing partitions

roughly equivalent parts. Although the error becomes arbitrary small, there is no way of getting rid of it. *At this point Savage uses the σ-algebra assumption, he puts*:

$$B_1 = \bigcup_n C_n \qquad \text{and} \qquad B_2 = \Big(\bigcup_n D_n\Big) \cup \Big(\bigcap_n G_n\Big). \tag{2.6}$$

Remark 2.14. The rest of Savage's proof is not relevant to our work. For the sake of completeness, here is a short account of it. B_1, B_2 form a partition of B, and $\bigcap_n G_n \equiv \varnothing$. Assuming P6$'$, one can show that $B_1 \equiv B_2$; but Savage does not use P6$'$ (a postulate that is introduced after Theorem 3), hence he only deduces that B_1 and B_2 are what he calls "almost equivalent" — one of the concepts he used at the time, which we need not go into. By iterating this division he proves that, for every n, every non-null event can be partitioned into 2^n almost equivalent events. At an earlier stage (Part 4) he states that every partition of S into almost equivalent events is almost uniform. Hence, there are almost uniform n-partitions of S for arbitrary large ns. This together with Theorem 2.11 proves the existence of the required numeric probability.

We eliminate the σ-algebra assumption by avoiding the construction (2.6). We develop, instead, a technique of using trees, which generates big partitions, and many "error parts", which can be treated simultaneously. We use it in order to get almost uniform partitions.

2.2 Eliminating the σ-Algebra Assumption by Using Tripartition Trees

So far, trying to follow faithfully the historical development of Savage's system, we presupposed (fine) rather than P6′. Otherwise the proof will be burdened by various small details, and we prefer to avoid this.[15] From now on, in order to simplify our proofs, we shall presuppose P6′.[16]

First, let us give the 3-partitions that figure in Savage's construction a more suggestive form, suitable for our purposes:

Definition 2.15 (Tripartition)**.** A *Savage tripartition* or, for short, a *tripartition* of a non-null event, B, is an ordered triple (C, E, D) of disjoint events such that:

i. $B = C \cup E \cup D$

ii. $C, D \succ \varnothing$,

iii. $C \cup E \succeq D$ and $E \cup D \succeq C$.

We refer to E as the *error-part*, or simply *error*, and to C and D as the *regular parts.*

We allow E to be a null-set, i.e., $E \equiv \varnothing$, including $E = \varnothing$. The latter constitutes the extreme case of a tripartition, where there the error is $\varnothing$. In diagrams, $\varnothing$ serves in this case a marker that separates the two parts.[17]

2.2.1 Tripartition Trees

Recall that a binary partition tree is a rooted ordered tree whose nodes are sets, such that each node that is not a leaf has two children that form a 2-partition of it.

By analogy, a *tripartition* tree is a rooted ordered tree such that: (1) The nodes are sets, which are referred to as *parts*, and they are classified into *regular parts*, and *error parts.* (2) The root is a regular part. (3) Every regular part that is not a leaf has three children that constitute a tripartition of it. (4) error-parts have no children.

Note. No set can occur twice in a partition tree. Hence we can simplify the structure by identifying the nodes with the sets; we do not have to construe it as a *labeled* tree. Later, in special occasions, the empty set can be an error-part and it can occur more than once among the leaves of the tree. This should not cause any confusion.

Figure 2.2 provides an illustration of a top-down tripartition, written top down, in which the root is the event A, and the error-parts are shaded.

[15]Under P6′, $E \equiv \varnothing$ implies $A \cup E \equiv A$; if only (fine) is assumed this need not hold , but it is still true that $A \cup E$ can be made arbitrary small, by making A arbitrary small.

[16]Our result still holds if we presuppose (fine) only, provided that the unique numeric probability is claimed to *almost represent*, rather than represent, the qualitative one. See (2.4) in Section 2.1.2 and the discussion there.

[17] Under the P6′ the case $E \equiv \varnothing$ can, for all purposes, be assimilated to the case $E = \varnothing$, because we can add E to one of the regular parts, say C, and $C \cup E \equiv C$. But under (fine) non-empty null-sets cannot be eliminated in this way.

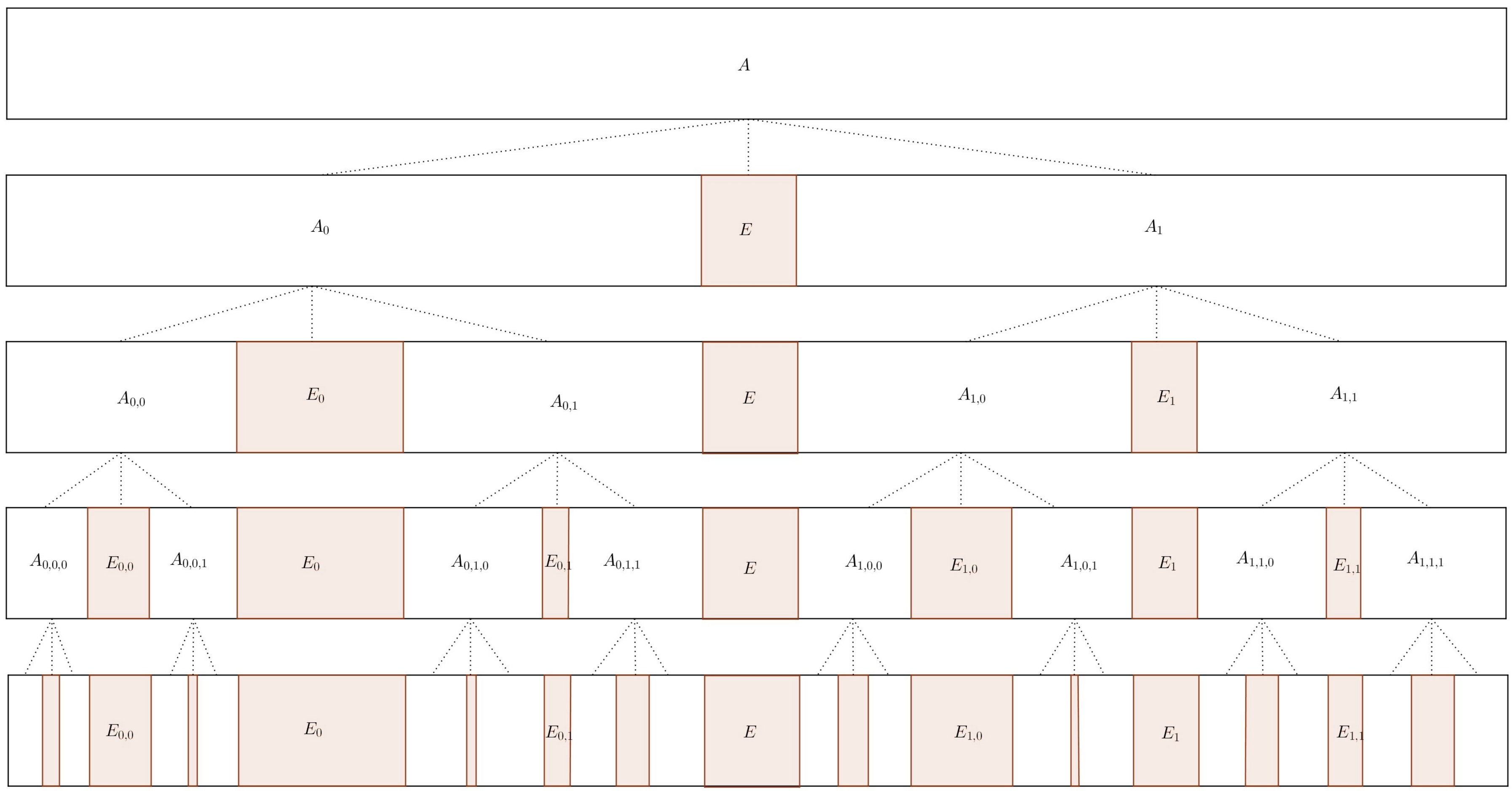

Figure 2.2: A tripartition tree $\mathcal{T}$ of A

Additional Concepts, Terminology, and Notations

1. The *levels* of a tripartition tree are defined as follows: (i) level 0 contains the root; (ii) level $n+1$ contains all the children of the regular nodes on level n; (iii) if level $n+1$ contains regular nodes, then it contains all error nodes on level n.

2. Note that this means that, once an error-part appears on a certain level it keep reappearing on all higher levels that contain regular nodes.

3. A tripartition tree is *uniform* if all the regular nodes that are leaves are on the same level. From now on we assume that the tripartition trees are uniform, unless indicated otherwise.

4. The *height* of a finite tree is n , where n is the level of the leaves that are regular nodes. If the tree is infinite its height is ∞.

5. A *subtree* of a tree, $\mathcal{T}$, is a tree consisting of some regular node (the root of the subtree) $\mathcal{T}$ and all its descendants.

6. The *truncation* of a tree, $\mathcal{T}$, at level m, is the tree consisting of all the nodes of $\mathcal{T}$ whose level is $\leq m$. (Note that if $m \geq$ height of $\mathcal{T}$, then truncation at level m leave is the same as $\mathcal{T}$.)

7. Strictly speaking, the root by itself does not constitute a tripartition tree. But there is no harm in regarding it as the truncation at the 0 level, or as a tree of height 0.

Remark 2.16. (1) An *ordered* tree is one in which the children of any node are ordered (an assignment, which assigns to every node an ordering of its children, is included in the structure). Sometimes the trees must be ordered, e.g., when they are used to model syntactic structures of sentences. But sometimes an ordering is imposed for convenience; it makes for an easier location of nodes and for a useful two-dimensional representation. In our case, the ordering makes it possible to locate the error-parts by their middle positions in the triple.[18]

(2) The *main* error part of a tree is the error part on level 1.

(3) It is easily seen that on level k there are 2^k regular parts and $2^k - 1$ error-parts. We use binary strings of length k to index the regular parts, and binary strings of length $k-1$ to index the error-parts, except the the main error-part. Figure 2.1 shows how this is done. The main error-part of that tree is E. We can regard the index of E as the empty binary sequence.

[18] Yet, the left/right distinction of the regular parts is not needed. Formally, we can take any regular part, B, which is not a leaf, and switch around the two regular parts that appear in its tripartition: from $B_l\ B\ B_r$ to $A_r\ B\ A_l$, switching at the same time the subtrees that are rooted in A_l and A_r. The switch can be obtained by rotating (in a 3-dimensional space) the two subtrees. Such a switch can be considered an"automorphism" of the structure: Our tripartition trees can be viewed as ordered trees, "divided" by the equivalence that is determined by the automorphism group generated by these rotations. All the claims that we prove in the sequel hold under this transformation group.

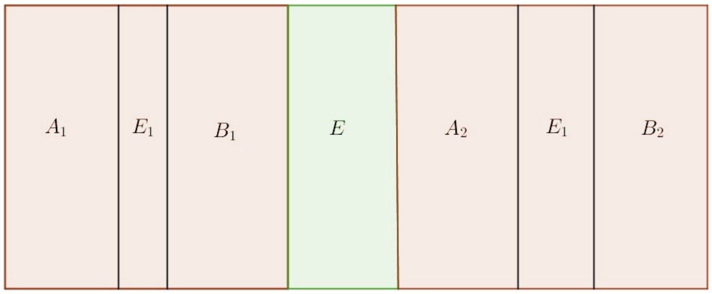

Figure 2.3: Claim in the proof of Theorem 2.18

(4) We let $\mathcal{T}$ range over tripartitions trees, and $\mathcal{T}_A$ — over tripartition trees of A. We put $\mathcal{T} = \mathcal{T}_A$ in order to say that $\mathcal{T}$ is a tripartition tree of A . To indicate the regular and error parts we put: $\mathcal{T}_A = (A_\sigma, E_\sigma)$, where σ ranges over the binary sequences (it is understood that the subscript of E ranges over sequences of length smaller by 1 then the subscript of A.) To indicate also the height k, we put: $\mathcal{T}_{A,k} = (A_\sigma, E_\sigma)_k$. Various parameters will be omitted if they are understood from the context.

Definition 2.17 (Total error)**.** The *total error* of a tree $\mathcal{T}$, denoted $E(\mathcal{T})$, is the union of all error-parts of $\mathcal{T}$. That is to say, if $\mathcal{T} = \mathcal{T}_A = (A_\sigma, E_\sigma)$, then $E(\mathcal{T}) =_{\text{Df}} \bigcup_\sigma E_\sigma$.

If $\mathcal{T}$ is of height k then $E(\mathcal{T})$ is the union of all error-parts on the k-level of $\mathcal{T}$. This is obvious, given that all error-parts of level j, where $j < k$, reappear on level $j+1$. For the same reason, if $j < k$, then the total error of the truncated tree at level j is the union of all error-parts on level j.

Now, recall that a Savage tripartition (C, E, D) has the property that $C \cup E \succeq D$ and $C \preceq E \cup D$ (cf. Definition 2.15). This property generalizes to tripartition trees:

Theorem 2.18. Let $\mathcal{T}_A$ be a partition tree of A of height k, then, for any regular parts $A_\sigma, A_{\sigma'}$ on the kth level, the following holds;

$$\begin{aligned} A_\sigma \cup E(\mathcal{T}_A) &\succeq A_{\sigma'} \\ A_{\sigma'} \cup E(\mathcal{T}_A) &\succeq A_\sigma \end{aligned} \tag{2.7}$$

Proof. We prove the theorem by induction on k. For $k = 1$ the claim holds since (A_0, E, A_1) is just a Savage tripartition. For $k > 1$, let $\mathcal{T}^*_A = (A_\sigma, E_\sigma)_{k-1}$ be the truncated tree consisting of the first $k-1$ levels of $\mathcal{T}_A$. By the inductive hypothesis,, for all regular parts $A_\tau, A_{\tau'}$ on the $k-1$ level of $\mathcal{T}^*_A$,

$$\begin{aligned} A_\tau \cup E(\mathcal{T}^*_A) &\succeq A_{\tau'} \\ A_{\tau'} \cup E(\mathcal{T}^*_A) &\succeq A_\tau \end{aligned} \tag{2.8}$$

The rest of the proof relies on the following claim.

Claim. Assume that the following holds, as illustrated in Figure 2.3:

$$\begin{aligned} A_1 \cup E_1 \succeq B_1 \qquad B_1 \cup E_1 \succeq A_1 \\ A_2 \cup E_2 \succeq B_2 \qquad B_2 \cup E_2 \succeq A_2 \end{aligned} \tag{2.9}$$

and

$$\begin{aligned}\Big(A_1 \cup E_1 \cup B_1\Big) \cup E &\succeq \Big(A_2 \cup E_2 \cup B_2\Big)\\ \Big(B_2 \cup E_2 \cup A_2\Big) \cup E &\succeq \Big(A_1 \cup E_1 \cup B_1\Big)\end{aligned} \tag{2.10}$$

Then $C_1 \cup (E_1 \cup E \cup E_2) \succeq C_2$ where C_1 is either A_1 or B_1 and C_2 is A_2 or B_2.

Proof of Claim. WLOG, it is sufficient to show this for the case $C_1 = A_1$ and $C_2 = A_2$. Thus, we have to prove:

$$A_1 \cup (E_1 \cup E \cup E_2) \succeq A_2. \tag{2.11}$$

Now, consider the following two cases:

(1) If $B_1 \succeq A_2$, then we have

$$(A_1 \cup E_1) \cup E \cup E_2 \succeq B_1 \cup E \cup E_2 \succeq A_2 \cup E \cup E_2 \succeq A_2.$$

(2) Otherwise $B_1 \prec A_2$, we show that the claim also holds under this assumption. Suppose, to the contrary, that (2.11) fails, that is, $A_2 \succ A_1 \cup (E_1 \cup E \cup E_2)$. Since $A_1, E_1, E, E_2, A_2, B_2$ are mutually exclusive, we have, via (2.3),

$$\begin{aligned}A_2 \cup B_2 &\succ A_1 \cup (E_1 \cup E \cup E_2) \cup B_2\\ &\succeq A_1 \cup E_1 \cup E \cup A_2\\ &\succ A_1 \cup E_1 \cup E \cup B_1.\end{aligned}$$

The second inequality holds because $E_2 \cup B_2 \succeq A_2$ in (2.9) and the third holds because the assumption in this case is that $A_2 \succ B_1$. But, again from (2.9), we have that $A_1 \cup E_1 \cup E \cup B_1 \succeq A_2 \cup E_2 \cup B_2 \succeq A_2 \cup B_2$, a contradiction. Hence, it must be that (2.11) holds.

By symmetry, other cases hold as well. This completes the proof of the Claim.

Getting back to the proof of the theorem, assume WLOG that in (2.7) A_σ is to the left of A'_σ. Now each of them is a regular part of a tripartition of a regular part on level $k-1$. Consider the case in which A_σ appears in a tripartition of the form $(A_\sigma, E_\lambda, B_\sigma)$ and A'_σ appears in a tripartition of the form $(B_{\sigma'}, E_{\lambda'}, A_{\sigma'})$. There are other possible cases, but the argument in each case is of the same kind. We get:

$$\begin{aligned}A_\sigma \cup E_\lambda \succeq B_\sigma \qquad & B_\sigma \cup E_\lambda \succeq A_\sigma\\ A_{\sigma'} \cup E_{\lambda'} \succeq B_{\sigma'} \qquad & B_{\sigma'} \cup E_{\lambda'} \succeq A_{\sigma'}.\end{aligned} \tag{2.12}$$

Since $A_\sigma \cup E_\lambda \cup B_\sigma$ and $A_{\sigma'} \cup E_{\lambda'} \cup B_{\sigma'}$ are regular parts on the $k-1$ level of $\mathcal{T}_A$, the inductive hypothesis, (2.8) implies:

$$\begin{aligned}\Big(A_\sigma \cup E_\lambda \cup B_\sigma\Big) \cup E(\mathcal{T}^*_A) &\succeq \Big(A_{\sigma'} \cup E_{\lambda'} \cup B_{\sigma'}\Big)\\ \Big(A_{\sigma'} \cup E_{\lambda'} \cup B_{\sigma'}\Big) \cup E(\mathcal{T}^*_A) &\succeq \Big(A_\sigma \cup E_\lambda \cup B_\sigma\Big).\end{aligned} \tag{2.13}$$

Clearly, (2.12) and (2.13) are a substitution variant of (2.9) and (2.10). Therefore the Claim implies:

$$\begin{aligned} A_\sigma \cup \big(E(\mathcal{T}^*) \cup E_\lambda \cup E_{\lambda'}\big) \succeq A_{\sigma'} \\ A_{\sigma'} \cup \big(E(\mathcal{T}^*) \cup E_\lambda \cup E_{\lambda'}\big) \succeq A_\sigma. \end{aligned} \tag{2.14}$$

Since $E(\mathcal{T})$ is disjoint from A_σ and $A_{\sigma'}$ and $E(\mathcal{T}^*) \cup E_\lambda \cup E_{\lambda'} \subseteq E(\mathcal{T})$, we get (2.7). □

2.2.2 The Error Reduction Method for Trees

Note that trees that have the same height are structurally isomorphic and there is a unique one-to-one correlation that correlates the parts of one with the parts of the other. We have adopted a notation that makes clear, for each part in one tree, the corresponding part in the other tree. This also holds if one tree is a truncation of the other, The indexing of the regular parts and the error parts in the truncated tree is the same as in the whole tree.

Definition 2.19 (Error Reduction Tree)**.** Given a tree, $\mathcal{T}_A = (A_\sigma, E_\sigma)_k$, an *error-reduction of* $\mathcal{T}$ is a tree with the same root and the same height $\mathcal{T}'_A = (A'_\sigma, E'_\sigma)_k$, such that for every σ, $A_\sigma \subseteq A'_\sigma$. We shall also say in that case that $\mathcal{T}'$ is *obtained from* $\mathcal{T}$ *by error reduction.*

Remark 2.20. (1) Obviously, if all regular parts increase the total error must decrease, that is: For all σ, $A_\sigma \subseteq A'_\sigma \implies E(\mathcal{T}') \subseteq E(\mathcal{T})$. Thus, the term 'error-reduction' is justified. The reverse implication is of course false in general. The crucial property of error-reducing is that, in the reduction of the total error, *every* regular part (weakly) increases as a set.

(2) The reduction of $E(\mathcal{T})$ is in the weak sense: that is, $E(\mathcal{T}') \subseteq E(\mathcal{T})$. The strong sense can be obtained by adding the condition $E(\mathcal{T}') \prec E(\mathcal{T})$. But, in view of our main result, we shall not need it

(3) Error reductions of tripartitions is the simplest case of error reduction of trees: each of the two regular parts weakly increases and the error part weakly decreases. Note that this is the error reduction in trees of height 1.

(4) It is easily seen that if $\mathcal{T}'$ is an error-reduction of $\mathcal{T}$ and $\mathcal{T}''$ is an error-reduction of $\mathcal{T}'$, then $\mathcal{T}''$ is an error-reduction of $\mathcal{T}$.

The proof of our central result is that, given any tripartition tree, there is an error-reduction of it in which the total error is arbitrarily small; that is, for every non-null set F, there is an error-reduction tree of total error $\preceq F$. The proof is based on a certain operation on tripartition trees, which is defined as follows.

Definition 2.21 (Mixed Sum)**.** Let $\mathcal{T}_A = (A_\sigma, E_\sigma)_\sigma$ and $\mathcal{T}'_{A'} = (A'_\sigma, E'_\sigma)_\sigma$ be two tripartition trees of two disjoint events (i.e., $A \cap A' = \varnothing$), of the same height, k. Then the *mixed sum* of $\mathcal{T}_A$ and $\mathcal{T}'_{A'}$, denoted $\mathcal{T}_A \oplus \mathcal{T}'_{A'}$, is the tree of height k, defined by:

$$\mathcal{T}_A \oplus \mathcal{T}'_{A'} = (A_\sigma \cup A'_\sigma,\ E_\sigma \cup E'_\sigma)_\sigma \tag{2.15}$$

The notation $\mathcal{T}_A \oplus \mathcal{T}'_{A'}$ is always used under the assumption that A and A' are disjoint and the trees are of the same height.

Lemma 2.22. 1. $\mathcal{T}_A \oplus \mathcal{T}'_{A'}$ is a tripartition tree of $A \cup A'$ whose total error is $E(\mathcal{T}_A) \cup E(\mathcal{T}'_{A'})$.

2. If $\mathcal{T}^*_A$ and $\mathcal{T}^+_{A'}$ are, respectively, error reductions of $\mathcal{T}_A$ and $\mathcal{T}'_{A'}$, then $\mathcal{T}^*_A \oplus \mathcal{T}^+_{A'}$ is an error reduction of $\mathcal{T}_A \oplus \mathcal{T}'_{A'}$

Proof. The operation $\oplus$ consists in taking the union of every pair of corresponding parts, which belong to tripartitions of two given disjoint sets. Therefore, the first claim follow easily from the definitions of tripartition trees and the laws of qualitative probability (cf. Definition 2.2). For example, for every binary sequence, σ, of length $<$ height of the tree, we have $A_{\sigma,0} \cup E_\sigma \succeq A_{\sigma,1}$ and $A'_{\sigma,0} \cup E'_\sigma \succeq A'_{\sigma,1}$. In each inequality the sets are disjoint, and every set in the first inequality is disjoint from every set in the second inequality. Hence, by the axioms of qualitative probability we get:

$$\left(A_{\sigma,0} \cup A'_{\sigma,0}\right) \cup \left(E_\sigma \cup E'_\sigma\right) \succeq \left(A_{\sigma,1} \cup A'_{\sigma,1}\right)$$

The second claim follows as easily from the definition of error-reduction and the laws of Boolean algebra. □

Theorem 2.23 (Error Reduction)**.** For any tripartition tree $\mathcal{T}_A$ and any non-null event F, there is an error-reduction tripartition $\mathcal{T}^*_A$ such that $E(\mathcal{T}^*_A) \preceq F$.

Proof. We prove the theorem by induction on k, where $k =$ height of $\mathcal{T}_A$. If $k = 0$, then formally $\mathcal{T}_A$ consists of A only. Hence the base case is $k = 1$, and the only error part is on level 1. Let the tripartition on level 1 be (A_0, E, A_1). We now use a result that is implied by Fishburn's reconstruction of the proofs that Savage did not include in his book:[19]

Claim. Given any tripartition (C_0, E_0, D_0), there is a sequence of of tripartitions (C_n, E_n, D_n), $n = 1, 2, ...$ that constitute a Savage chain such that (C_1, E_1, D_1) is an error reduction of (C_0, E_0, D_0).

Applying this Claim, we first get an error reduction of (A_0, E, A_1), then continue to reduce the error via the Savage chain until, for some n, $E_n \preceq F$. This proves the base case.

Before proceeding, observe that, for any integer $m > 1$, every non-null event F can be partitioned into m disjoint non-null events. This is an easy consequence of (fine).[20] In what follows we use a representation of ordered partition trees of the form:

$$[\mathcal{T}_{B_1}, ..., \mathcal{T}_{B_m}]$$

where $m > 1$ and the B_is are disjoint non-null sets. This includes the possibility that some $\mathcal{T}_{B_i}$s are of height 0, in which case we can replace $\mathcal{T}_{B_i}$ by B_i. The root of the tree is the union of the B_is, the B_i's are its children, ordered as indicated by the indexing. The whole tree is not a tripartition tree but each of the m subtrees is. For example, $[B, B', \mathcal{T}_C, \mathcal{T}_D]$ denotes

[19] See the proof of C8 (and the claims that lead to it) in Fishburn (1970, p.195-198)

[20] Since $F \succ \varnothing$ there exists a non-null subset $F_1 \subseteq F$ such that $F \succ F_1 \succ \varnothing$. This is established by considering an n-partition $S = S_1 \cup ... \cup S_n$ such that $S_i \prec F$ for all $i = 1, ..n$, and observing that there must be two different parts, say S_i, S_j, whose intersections with F are $\succ \varnothing$; otherwise, $F \preccurlyeq S_k$, for some k, which is a contradiction. Put $F_1 = F \cap S_i$ then F_1 and $F - F_1$ are non-null, and we can apply the same procedure to $F - F_1$, and so on.

a partition tree in which (B, B', C, D) is a 4-partition of the root, the root being the union of these sets, B and B' are leaves, and C and D are roots of the tripartition trees $\mathcal{T}_C$ and $\mathcal{T}_D$.

Assume now that the induction hypothesis holds for k and let $\mathcal{T}_A$ be a tripartition tree of height $k+1$. Then, treating $\mathcal{T}_A$ is of the form:

$$(\mathcal{T}_{B_l},\ E,\ \mathcal{T}_{B_r})$$

where $\mathcal{T}_{B_l}$ and $\mathcal{T}_{B_r}$ are of height k. Next, partition the given F into 5 non-null events: F_1, F_2, F_3, F_4, F_5; as observed above this is always possible.

If E is a null set, then we apply the induction hypotheses to each of $\mathcal{T}_{B_l}$ and $\mathcal{T}_{B_r}$, get error-reductions in which the total errors are, respectively, less-that-or-equal-to F_1 and F_5, and we are done. Otherwise we proceed as follows.

Using Savage's theorem, we can replace E by C_l, E^*, C_r, where $E^* \preceq F_3$. Ignoring for the moment the role of E^* as an error part, we get:

$$[\mathcal{T}_{B_l},\ C_l,\ E^*,\ C_r,\ \mathcal{T}_{B_r}]$$

Note that in this partition the root, which is A, is first partitioned into 5 events; B_l and B_r are roots of tripartition trees of height k, and C_l, E^*, and C_r are leaves. Using the induction hypothesis, get an error-reduction $\mathcal{T}^*_{B_l}$ of $\mathcal{T}_{B_l}$ and an error-reduction $\mathcal{T}^*_{B_r}$ of $\mathcal{T}_{B_r}$, such that $E(\mathcal{T}^*_{B_l}) \preceq F_1$, and $E(\mathcal{T}^*_{B_r}) \preceq F_5$. Get an arbitrary tripartition $\mathcal{T}_{C_l}$ of C_l, and an arbitrary tripartition $\mathcal{T}_{C_r}$ of C_r each of height k (every non-null set has a tripartition tree of any given height). Using again the inductive hypothesis, get error-reductions, $\mathcal{T}^*_{C_l}$ and $\mathcal{T}^*_{C_r}$, such that $E(\mathcal{T}^*_{C_l}) \preceq F_2$, and $E(\mathcal{T}^*_{C_l}) \preceq F_4$. This gives us the following partition of A:

$$[\mathcal{T}^*_{B_l},\ \mathcal{T}^*_{C_l},\ E^*,\ \mathcal{T}^*_{C_r},\ \mathcal{T}^*_{B_r}]$$

Now, put $\mathcal{T}_{A_0} = \mathcal{T}^*_{B_l} \oplus \mathcal{T}^*_{C_l}$ and $\mathcal{T}_{A_1} = \mathcal{T}^*_{B_r} \oplus \mathcal{T}^*_{C_r}$, then

$$(\mathcal{T}_{A_0},\ E^*,\ \mathcal{T}_{A_1})$$

is a tripartition tree of A of height $k+1$. Call it $\mathcal{T}^*_A$. By Lemma 2.22, $E(\mathcal{T}_{A_0}) \preceq F_1 \cup F_2$ and $E(\mathcal{T}_{A_1}) \preceq F_4 \cup F_5$. Since $E^* \preceq F_3$, together we get: $E(\mathcal{T}^*_A) \preceq F$. □

Note. In a way, this theorem generalizes the construction of monotonically decreasing sequence of error-parts in Theorem 2.12. But, instead of reducing a single error-part (the shaded areas in Figure 2.1), the method we use reduces simultaneously *all* error-parts in a tripartition tree.

2.2.3 Almost Uniform Partition

Recall that a partition $\{P_i\}_{i=1}^n$ of a non-null event A is *almost uniform* if the union of any r members of the partition is no more probable than the union of any $r+1$ members. In Theorem 2.11 we rephrased a result by Savage, which claims that if, for infinitely many ns, there are almost uniform n-partitions of S, then there is a unique numeric probability that almost represents the qualitative one. We noted that Savage's proof requires no further assumptions regarding the qualitative probability, and that if we assume P6′ then the probability (fully) represents the qualitative one (cf. Remark 2.8 above). It therefore remains to show that, using Theorem 2.23, which we have just proved, we can derive the existence of almost uniform partitions of S of arbitrary large size (number of parts). The derivation is based on repeated error reduction in tripartition trees.

Definition 2.24. Given $C \succ \varnothing$, let us say that $B \ll \frac{1}{n}C$ if there is a sequence $C_1, C_2, ...C_n$, of n disjoint subsets of C, such that $C_1 \preceq C_2 \preceq \cdots \preceq C_n$ and $B \preceq C_1$.

The following are some simple intuitive properties of $\ll$. The first two are immediate from the definition, and in the sequel we shall need only the first.

Lemma 2.25.
1. If $B \ll \frac{1}{n}C$, and if $A \preceq B \ll \frac{1}{n}C \subseteq D$ then $A \ll \frac{1}{n}D$.[21]
2. If $B \ll \frac{1}{n}C$ then $B \ll \frac{1}{m}C$ for all $m < n$.
3. For any $C, D \succ \varnothing$, there exists n such that, for all B,

$$B \ll \frac{1}{n}C \implies B \preceq D. \tag{2.16}$$

Lemma 2.26. Let $\mathcal{T} = (A_\sigma, E_\sigma)$ be a tripartition tree of height k, then, given any n and any regular part A_σ on the kth level of $\mathcal{T}$, there is an error reduction $\mathcal{T}'$ of $\mathcal{T}$ such that

$$E(\mathcal{T}'_A) \ll \frac{1}{n}A'_\sigma \tag{2.17}$$

where A'_σ is the part that corresponds to A_σ under the structural isomorphism of the two trees.

Proof. Fix A_σ and let $\{C_i\}_{i=1}^n$ be a disjoint sequence of events contained in it as subsets, such that $C_1 \preceq C_2 \preceq \ldots \preceq C_n$. Applying the Error Reduction Theorem 2.23, get a tree $\mathcal{T}'$ such that $E(\mathcal{T}') \preceq C_1$. Consequently, $E(\mathcal{T}'_A) \ll \frac{1}{n}A_\sigma$. Since the parts are disjoint and under the error reduction each regular part in $\mathcal{T}$ is a subset of its corresponding part in $\mathcal{T}'$, A'_σ is the unique part containing A_σ as a subset, the required result then follows. □

Lemma 2.27. Given any tripartition tree $\mathcal{T} = (A_\sigma, E_\sigma)$ of height k and given any n, there is an error reduction $\mathcal{T}' = (A'_\sigma, E'_\sigma)$ of $\mathcal{T}$ such that, for every regular part A'_σ on the kth level, $E(\mathcal{T}') \ll \frac{1}{n}A'_\sigma$.

Proof. Apply Lemma 2.26 repeatedly 2^k times, as σ ranges over all the binary sequences of length k. Since the regular parts can only expand and the total error can only contract, we get at the end an error reduction, $\mathcal{T}'$, such that $E(\mathcal{T}') \ll \frac{1}{n}A'_\sigma$, for all σ. □

[21]Note however that from $B \ll \frac{1}{n}C$ and $C \preceq D$ we cannot infer $B \ll \frac{1}{n}D$. The inference is true if we assume $C \prec D$; this can be shown by using the numeric probability that represents the qualitative one — whose existence we are about to prove. There seems to be no easier way of showing it.

Remark 2.28. Say that a property of a tripartition tree is *persistent* if whenever it holds for a tree $\mathcal{T}$ it also holds for every error reductionof $\mathcal{T}$. Then the property proved in Lemma 2.27 is persistent because, by our definition of error reducing refinement, all the regular parts can only expand and the total errors can only shrink. This implies that, for all refinements $\mathcal{T}''$ of $\mathcal{T}'$, we have $E(\mathcal{T}'') \ll \frac{1}{n} A''_\sigma$.

Theorem 2.29. Let $\mathcal{T}$ be a tripartition tree of height k, then there is an error reduction$\mathcal{T}'$ of $\mathcal{T}$ such that if Ξ_1 and Ξ_2 are any two sets of regular parts of the kth level of $\mathcal{T}'$ that are of equal cardinality r $(< 2^{k-1})$ and if A'_τ is any regular part on the kth level that is not in Ξ_1 or Ξ_2, then we have

$$\bigcup_{A'_\sigma \in \Xi_1} A'_\sigma \cup E(\mathcal{T}') \preceq \bigcup_{A'_\sigma \in \Xi_2} A'_\sigma \cup A'_\tau. \tag{2.18}$$

Proof. WLOG, assume that $\Xi_1 \cap \Xi_2 = \varnothing$ (otherwise, common members can be canceled out). Apply Lemma 2.27 for the case where $n = 2^{k-1}$, then we get a reduction tree $\mathcal{T}'$ of $\mathcal{T}$ such that $E(\mathcal{T}') \ll \frac{1}{2^{k-1}} A'_\sigma$ for all regular parts A'_σ on the kth level of $\mathcal{T}'$. Let Ξ_1 and Ξ_2 and A'_τ be as in the statement of the theorem, then we have

$$E(\mathcal{T}') \ll \frac{1}{2^{k-1}} A'_\tau.$$

By Definition 2.24, this means there is a disjoint sequence $\{C_i\}$ of subsets of A'_τ of length 2^{k-1} such that

$$\begin{gathered} E(\mathcal{T}') \preceq C_1 \preceq C_2 \preceq \cdots \preceq C_r \preceq C_{r+1} \preceq \cdots \preceq C_{2^{k-1}} \\ \bigcup_{i=1}^{2^{k-1}} C_i \subseteq A'_\tau \end{gathered} \tag{2.19}$$

where r is the cardinality of Ξ_1 and Ξ_2 and $r < 2^{k-1}$.

Now, let $A_1, A_2 \ldots, A_r$ and $B_1, B_2 \ldots, B_r$ be enumerations of members of Ξ_1 and Ξ_2, respectively. Obviously, we have $E(\mathcal{T}') \preceq A_i$ and $E(\mathcal{T}') \preceq B_i$ for all $(i = 1, \ldots, r)$. And since $E(\mathcal{T}') \preceq C_1$, apply Theorem 2.18, we get, via (2.19), that

$$A_i \preceq B_i \cup E(\mathcal{T}') \preceq B_i \cup C_i \qquad \text{for all } i = 1, \ldots, r.$$

Further, since all parts considered here are disjoint, we get $\bigcup_{i=1}^r A_i \preceq \bigcup_{i=1}^r B_i \cup \bigcup_{i=1}^r C_i$, that is,

$$\bigcup_{A'_\sigma \in \Xi_1} A'_\sigma \preceq \bigcup_{A'_\sigma \in \Xi_2} A'_\sigma \cup \bigcup_{i=1}^r C_i. \tag{2.20}$$

On the other hand, from (2.19), we have $E(\mathcal{T}') \preceq C_{r+1} \preceq \cdots \preceq C_t$. Again, since all parts are disjoint, then we get

$$\begin{aligned} \bigcup_{A'_\sigma \in \Xi_1} A'_\sigma \cup E(\mathcal{T}') &\preceq \bigcup_{A'_\sigma \in \Xi_2} A'_\sigma \cup \bigcup_{i=1}^r C_i \cup C_{r+1} \\ &\preceq \bigcup_{A'_\sigma \in \Xi_2} A'_\sigma \cup \bigcup_{i=1}^{2^{k-1}} C_i \preceq \bigcup_{A'_\sigma \in \Xi_2} A'_\sigma \cup A'_\tau \end{aligned} \tag{2.21}$$

which is what we want. ∎

Remark 2.30. Let $\mathcal{T}$ be any tripartition tree of height k, the proceeding theorem shows that there exists an error reduction tree $\mathcal{T}'$ of $\mathcal{T}$ for which (2.18) holds. Suppose that Ξ is a union of r many regular parts on the kth level of $\mathcal{T}'$, and Ξ' is a union of $r+1$ many regular parts of the same kind, then we have

$$\bigcup_{A'_\sigma \in \Xi} A'_\sigma \cup E(\mathcal{T}') \preceq \bigcup_{A'_\sigma \in \Xi'} A'_\sigma. \tag{2.22}$$

It is now easily seen that an almost uniform partition can be formed by joining $E(\mathcal{T}')$ to any regular part of the highest level of $\mathcal{T}'$.

As we have remarked earlier, once the existence of an almost uniform partition is proved, numeric representation of $\succeq$ can be established by following Savage's methods (cf. Theorem 2.11 above). This completes our construction of numeric probability without the σ-algebra assumption by using tripartition trees.

2.2.4 The Proof of the (‡) Condition

Next we demonstrate that the (‡) condition holds. As we shall show in Section 3, this property will play a crucial role in defining utilities for simple acts without the σ-algebra assumption.

Theorem 2.31. Let μ be the probability that represents the qualitative probability $\succeq$. Assume that P6$'$ holds.[22] Then, for every non-null event, A, every $\rho \in (0,1)$ and every $\epsilon > 0$ there exists an event $B \subseteq A$, such that $(\rho - \epsilon) \cdot \mu(A) \leq \mu(B) \leq \rho \cdot \mu(A)$.

Proof. As shown by Savage, there is a Savage chain for A, that is, an infinite sequence of 3-partitions of A: $(A'_n\ E_n\ A''_n)_n$, $n = 1, 2, ...$ such that:

(i) $A'_n \cup E_n \succeq A''_n$ and $A'_n \cup E_n \succeq A''_n$

(ii) $(A'_{n+1} \supseteq A'_n)$, $(A''_{n+1} \supseteq A''_n)$, hence $E_{n+1} \subseteq E_n$

(iii) $E_n - E_{n+1} \succeq E_{n+1}$.

Consequently we get:

(1) $\mu(A'_n) + \mu(E_n) \geq \mu(A''_n) \qquad \mu(A''_n) + \mu(E_n) \geq \mu(A'_n)$, which implies:

  (a) $|\mu(A'_n) - \mu(A''_n)| \leq \mu(E_n)$.

(2) $\mu(E_{n+1}) \leq (1/2) \cdot \mu(E_n)$, which implies:

  (b) $\mu(E_n) \leq (1/2)^{n-1}$.

Since $\mu(A) = \mu(A'_n) + \mu(E_n) + \mu(A''_n)$, we get from (a) and (b):

$$\mu(A'_n) \longrightarrow 1/2 \cdot \mu(A), \qquad \mu(A''_n) \longrightarrow 1/2 \cdot \mu(A).$$

Since both A'_n and A''_n are monotonically increasing as sets, $\mu(A'_n)$ and $\mu(A''_n)$ are monotonically increasing. Consequently, we get: $\mu(A'_n) \leq 1/2 \cdot \mu(A)$ and $\mu(A''_n) \leq 1/2 \cdot \mu(A)$. All these imply the following claim:

[22] Actually (fine) is sufficient.

Claim 1. Let A be a non-null set. Then, for every $\epsilon > 0$, there are two disjoint subsets of A, A_0 and A_1, such that, for $i = 0, 1$:

$$1/2 \cdot \mu(A) - \epsilon \quad \leq \quad \mu(A_i) \quad \leq \quad 1/2 \cdot \mu(A).$$

Call such a partition an ϵ-*bipartition* of A. Call ϵ the error-margin of the bipartition. We can now apply such a bipartition to each of the parts, and so on. By "applying the procedure" we mean applying it to all the non-null minimal sets that were obtained at the previous stages (the inductive definition should be obvious).

Claim 2. Let A be any non-null set. Then for every $k > 1$ and every $\epsilon > 0$, there are 2^k disjoint subsets of A, $A_i, i = 1, ..., 2^k$, such that:

$$1/2^k \cdot \mu(A) - \epsilon \quad \leq \quad \mu(A_i) \quad \leq \quad 1/2^k \cdot \mu(A).$$

(This claim is proved by considering k applications of the procedure above, where the error-margin is ϵ/k.)

Now, note that since Claim 2 is made for any $\epsilon > 0$, and any $k > 1$, we can replace ϵ by $\epsilon/2^k \cdot \mu(A)$. Thus, the following holds:

(+) For every $\epsilon > 0$, $k > 1$, there are 2^k disjoint subsets, A_i, of A, such that:

$$1/2^k \cdot \mu(A) - \epsilon/2^k \cdot \mu(A) \quad \leq \quad \mu(A_i) \quad \leq \quad 1/2^k \cdot \mu(A)$$

The following is a different form of (+)

($*$) For every $\epsilon > 0$, $k > 1$, there are 2^k disjoint subsets, A_i, of A, such that:

$$\mu(A_i) \in \left[1/2^k \cdot (\mu(A) - \epsilon), \quad 1/2^k \cdot \mu(A)\right]$$

Now the (‡) condition can be put in the form

($**$) Fix any non-null set A. Then for every $\rho < 1$, and any $\epsilon' > 0$, there is a set $B \subseteq A$, for which $\mu(B) \in [(1 - \epsilon') \cdot \mu(A), \mu(A)]$

All the subsets that are generated in the process above are subsets of A. (A plays the role of the "universe", except that its probability can be < 1.) It is not difficult to see that, depending on the given ρ and ϵ', we can chose (in ($*$)) our ϵ small enough and k large enough, so as to get sufficiently many disjoint sets whose conditional probabilities $\mu(\cdot|A)$ sum up to a number in the interval $[(1 - \epsilon'), 1]$. This concludes the proof of (‡). □

Remark 2.32. It's worth repeating that (‡) does not rely on the σ-algebra assumption, but (†) does.

2.3 Countable Models

The σ-algebra assumption implies that the Boolean algebra of events has at least the cardinality of the continuum. Its elimination makes it possible to use a countable Boolean algebra. All that is needed is a qualitative probability, $\succeq$, defined over a countable Boolean algebra, which satisfies P6$'$. There are more than one way to do this. Here is a type of what we shall call *bottom up extension.* In what follows, a *qualitative probability space* is a system of the form $(S, \mathcal{B}, \succeq)$, where $\mathcal{B}$ is a Boolean algebra of subsets of S and $\succeq$ is qualitative probability defined over $\mathcal{B}$.

Definition 2.33. Let $(S, \mathcal{B}, \succeq)$ be a qualitative probability space. Then a normal *bottom up extension* of $(S, \mathcal{B}, \succeq)$ is a pair consisting of a qualitative probability $(S', \mathcal{B}', \succeq')$ and a mapping $h : S' \longrightarrow S$, of S' onto S, such that for every $A \in \mathcal{B}$, $h^{-1}(B) \in \mathcal{B}'$ and $A \succeq B \iff h^{-1}(A) \succeq' h^{-1}(B)$.

Remark 2.34. The extension is obtained by, so to speak, splitting the atoms, (the states in S) of the original algebra. This underlies the technique of getting models that satisfy P6$'$. In order to satisfy P6$'$ we have, given $A \succeq B$, to partition S into sufficiently fine parts, $P_i, i = 1, 2, ..., n$, such that $A \succeq P_i$ for all P_i. If we start with a finite Boolean algebra, the way to do it is to divide the atoms into smaller atoms. The intuitive idea is that our states do not reflect certain features of reality, and that, if we take into account such features, some states will split into smaller ones.

This picture should not imply that P6$'$, which is a technical condition, should be adopted. The intuitive justification of P6$'$, which has been pointed out by Savage, is different. But it can serve as a good background for repeated extensions.

We have shown that, starting from a finite qualitative probability space we can, by an infinite sequence of normal extensions get a countable space (that is, both S and $\mathcal{B}$ are countable) that satisfies P6$'$. We can also get models with other desired features.

Theorem 2.35 (Countable Model Theorem)**.** (1) Let $(S_0, \mathcal{B}_0, \succeq_0)$ be a finite qualitative probability space and that the qualitative probability is representable by some numeric probability. Then there is an infinite countable model, $(S, \mathcal{B}, \succeq)$, which forms together with a mapping, $h : S \to S_0$, a normal extension of $(S_0, \mathcal{B}_0, \succeq_0)$, and which satisfies P6$'$.

(2) Let Ξ be any countable subset of $(0, 1)$ and let μ be the numeric probability that represents $\succeq$ (which exists by our results). Then we can construct the model $(S, \mathcal{B}, \succeq)$ in such a way that $\mu(A) \notin \Xi$ for every $A \in \mathcal{B}$.

This theorem implies, for example, that for all n, no number of the form $1/n$, where $n > 1$, and no number of the form $(1/2)^n$, where $n > 0$, are among the values of μ. Now de Finetti and Koopman proposed axiom systems for subjective probability that included an axioms stating that there are partitions of S into n equal parts for arbitrary large ns. Our theorem shows that without the σ-algebra assumption this cannot be done. Savage found P6$'$ more intuitive than their axioms (and indeed it is), but was somewhat puzzled by the fact that it implies their axioms. Although it is more intuitive it appears to be stronger. Our last theorem solves this puzzle. It shows that without the σ-algebra assumptionit does not imply their axioms.

Remark 2.36. So far we have been dealing with the Boolean algebra only. But in order to state the results within the full perspective of Savage's system, we shall state them as results about decision models, that is, about systems of the form $(S, \mathbf{C}, \mathcal{A}, \succcurlyeq, \mathcal{B})$. This is done in the following theorem.

In what follows $f \circ g$ is the composition of the functions f and g, defined by $(f \circ g)(x) = f(g(x))$. It is used under the assumption that the domain of f includes the range of g.

Theorem 2.37. Let $(S, \mathbf{C}, \mathcal{A}, \succcurlyeq, \mathcal{B})$ be a decision model that satisfies P1-P5 (where P5 is interpreted as the existence of two non-equivalent constant acts, but without assuming CAA). Assume that S is finite and there is a probability over $\mathcal{B}$ that represents the qualitative probability. Then there is a Savage system, $(S^*, \mathbf{C}^*, \mathcal{A}^*, \succcurlyeq^*, \mathcal{B}^*)$, that satisfies P1-P6 and there is a function h that maps S^* onto S such that the following holds:

(i) S^* and $\mathcal{B}^*$ are countable.

(ii) for all $A \in \mathcal{B}$, $h^{-1}(A) \in \mathcal{B}^*$,

(iii) $\mathbf{C}^* = \mathbf{C}$,

(iv) $f \in \mathcal{A}^*$ iff $f \circ h \in \mathcal{A}$

(v) $f^* \succcurlyeq^* g^*$ iff $f \circ h \succcurlyeq g \circ h$.

The proofs of our last two theorems rely on repeated extension techniques that are used in set theory. In our case, at every stage we have to ensure that a particular instance of P6$'$ should be satisfied. As the model grows, there are more cases to take care of, but we can arrange these tasks so that after the infinite sequence of extensions all are taken care of. We shall not go into more detail here.

3 A Simpler Utility Function for Simple Acts

In a section on extension of utility to more general acts, Savage made the following remarks:

> The requirement that an act has only a finite number of consequences may seem, from a practical point of view, almost no requirement at all. To illustrate, the number of time intervals that might possibly be the duration of a human life can be regarded as finite, if you agree that the duration may as well be rounded to the nearest minute, or second, or microsecond, and that there is almost no possibility of its exceeding a thousand years. More generally, it is plausible that, no matter what set of consequences is envisaged, each consequence can be particularly identified with some element of a suitably chosen finite, though possibly enormous, subset. If that argument were valid, it could easily be extended to reach the conclusion that infinite sets are irrelevant to all practical affairs, and therefore to all parts of applied mathematics. (Savage, 1972, p.76-77)

Savage however goes on to say that it is of high mathematical interests to generalize utility function for simple acts to all acts. We however take that, as far as real world decision problems are concerned, simple acts are all it is needed. Especially in view of our discussion on conceptual realism, it is too high a demand to require our agent to be a professional mathematician.

Now, if the decision problem is presented as a finite set of simple acts, we can avoid both CAA and the σ-algebra assumption in defining utilities. The result is a more simpler definition of utility function for simple acts. In what follows we provide a sketch as to how this can be done.

Note that it is known, and anyone who follows Savage's derivation can easily check it, that in the Savage system all that is needed for defining the probabilities are two non-equivalent constant acts.[23] Instead of using CAA we posit 2CA, i.e., there are two non-equivalent acts. Assume that they are $\mathfrak{c}_0$ and $\mathfrak{c}_1$ and that their corresponding consequences are a_0 and a_1. Then Savage's notion of "more probable" in Definition 2.1 can be defined in terms of $\mathfrak{c}_0$ and $\mathfrak{c}_1$, hence (2.2) now takes the form

$$\mathfrak{c}_1|E + \mathfrak{c}_0|\overline{E} \succcurlyeq \mathfrak{c}_1|F + \mathfrak{c}_0|\overline{F}. \tag{3.1}$$

In what follows when we speak of subsets of S we assume, unless stated otherwise, that they are members of $\mathcal{B}$. We now assign utilities to simple acts as follows. First, to fix a utility scale, we define:

$$U[\mathfrak{c}_0] = 0 \qquad U[\mathfrak{c}_1] = 1$$

where $U[f]$ is the value, or utility of f. This means that $u(a_0) = 0$ and $u(a_1) = 1$, where $u(a)$ is the utility of the consequence a. It can be shown that, under P1-P6 and 2CA, the agent's preference relation among simple acts can be represented by a cardinal utility function U.

In what follows we first show how such a utility function can be constructed under the (†) condition, then we indicate how the (‡) can be used to replace (†), and hence get a proof without the σ-algebra assumption.

3.1 Constructing Utilities under the (†) Condition

Consider now any *feasible* consequences, i.e., any $a \in \mathbf{C}$ for which there exists an act g such that $g^{-1}(a)$ is not null. Let $A = g^{-1}(a)$ and let

$$\mathfrak{c}_A^* =_{\text{Df}} g|A + \mathfrak{c}_0|\overline{A}.$$

By definition, $\mathfrak{c}_A^*$ yields a if $s \in A$, status quo otherwise.

To define utilities, we compare $\mathfrak{c}_A^*$ with $\mathfrak{c}_0$. If $\mathfrak{c}_A^* \equiv \mathfrak{c}_0$, we put $u(a) = 0$. Otherwise there are three possibilities:

$$\text{(i)} \quad \mathfrak{c}_1 \succcurlyeq \mathfrak{c}_A^* \succ \mathfrak{c}_0 \quad \text{(ii)} \quad \mathfrak{c}_A^* \succ \mathfrak{c}_1 \quad \text{(iii)} \quad \mathfrak{c}_0 \succ \mathfrak{c}_A^*$$

[23] This observation is also noted in (Fishburn, 1981, p.161) where the author remarked that "[as far as obtaining a unique probability measure is concerned] Savage's $\mathscr{C}$ [i.e., the set of consequences] can contain as few as two consequences" (see also Fishburn, 1982, p.6). Fishburn (1970, §14.1-3) contains a clean exposition of Savage's proof of (2.1), and see especially §14.3 for an illustration of the role of P1-6 played in deriving numerical probability.

In each one of these possibilities, the utility of $\mathfrak{c}_A^*$ and that of a can be defined as follows. Let μ be the numeric probability derived under the (†) condition. Then for case (i), let

$$\rho = \sup\left\{\mu(B) \;\middle|\; B \subseteq A \text{ and } \mathfrak{c}_A^* \succcurlyeq \mathfrak{c}_1|B + \mathfrak{c}_0|\overline{B}\right\}. \tag{3.2}$$

Define

$$U[\mathfrak{c}_A^*] = \rho \qquad \text{and} \qquad u(a) = \frac{\rho}{\mu(A)}. \tag{3.3}$$

For case (ii), let $\rho = \sup\{\mu(B) \mid B \subseteq A \text{ and } \mathfrak{c}_1 \succcurlyeq \mathfrak{c}_A^*|B + \mathfrak{c}_0|\overline{B}\}$, define $U[\mathfrak{c}_A^*] = 1/\rho$ and $u(c) = 1/[\rho \cdot \mu(A)]$. Case (iii) in which the utility comes out negative is treated along similar lines and is left to the reader.

This assignment of utilities leads to a representation of the utility of any simple acts, f, as the expected utilities of the consequences, that appear as values of the act, where without loss of generality, we assume that each consequence a of f is a feasible consequence.

3.2 Constructing Utilities under the (‡) Condition

The construction above is given using the (†) condition. But without the the σ-algebra assumption we need to replace (†) with (‡) in our proof. Here we outline how this is done for the case $\mathfrak{c}_1 \succcurlyeq \mathfrak{c}_A^* \succ \mathfrak{c}_0$.

To this end, let μ be the numeric probability derived under the (‡) condition, i.e., through our tripartition tree method. Then, instead of (3.2), let

$$\rho = \sup\left\{\mu(B) \;\middle|\; \forall \epsilon > 0\; \exists B' \subseteq A \left[\mu(B) - \epsilon \leq \mu(B') \leq \mu(B) \text{ and } \mathfrak{c}_A^* \succcurlyeq \mathfrak{c}_1|B' + \mathfrak{c}_0|\overline{B'}\right]\right\}.$$

Define utilities of $\mathfrak{c}_A$ and a as in (3.3), then we are done.

A Ramsey's System

The following is an overview of Ramsey (1926). All page numbers refer to this publication.

Ramsey was guided by what he calls "the old-established way of measuring a person's belief", which is "to propose a bet, and see what are the lowest odds which he will accept" (p.170). He finds this method "fundamentally sound", but limited, due to the diminishing marginal utility of money (which means that the person's willing to bet may depend not only on the odds but also on the absolute sums that are staked). Moreover, the person may like or dislike the betting activity for its own stake, which can be a distorting factor. Ramsey's proposal is therefore based on the introduction of an abstract scale, which is supposed to measure true utilities, and on avoiding actual betting. Instead, the agent is supposed to have a preference relation, defined over gambles (called by Ramsey *options*), which are of the form:

$$\alpha \text{ if } p,\ \beta \text{ if } \neg p.$$

It means that the agent gets α if p is true, β otherwise; here p is a proposition and $\alpha, \beta, \ldots$ are entities that serve as abstract payoffs, due to their value for the agent. Among the gambles we have: α if p, α if not-p, which can be written as: "α for certain". (Note that this does *not* imply that the agent gets the same value in all possible worlds, because the possible world can carry by itself some additional value.) If neither of the two gambles G_1 and G_2 is preferred to the other, the agent is indifferent between them and they are considered to be equivalent.

Obviously, bets can be easily described as gambles of the above form. Ramsey does not use the more general form (of which he was certainly aware) α_1 if p_1, α_2 if p_2,..., α_n if p_n because, for his purpose, he can make do with $n = 2$. When he has to define conditional degrees of beliefs he uses gambles with $n = 3$.

Concerning propositions Ramsey tells us that he assumes Wittgenstein's theory, but remarks that probably some other theory could be used as well (p.177 footnote 1). As for $\alpha, \beta, \ldots$, his initial explanation is somewhat obscure.[24] But shortly afterwards it turns out that the values are attached to the possible worlds and that they can be conceived as equivalence classes of equi-preferable worlds:

> Let us call any set of all worlds equally preferable to a given world α value: we suppose that if a world α is preferable to β any world with the same value as α is preferable to any world with the same value as β and shall say that the value of α is greater than that of β. This relation 'greater than' orders values in a series. We shall use α henceforth both for the world and its value. (p.178)

Obviously, the preference relation should be transitive and this can be imposed by an axiom (or as a consequence of axioms). Also the equivalence relation mentioned above should have the expected properties.[25] We thus get an ordering of the values, which Ramsey denotes using the standard inequality signs; thus, $\alpha > \beta$ iff α *for for certain* is preferred to β *for for certain*. He also uses the Greek letters ambiguously; thus, if α *for certain* is equivalent to β *for certain*, this is expressed in the form: $\alpha = \beta$. All in all, the Greek letters range over an ordered set.

The main task now is to "convert" this ordered set into the set of reals, under their natural ordering. Ramsey takes his cue from the historical way, whereby real numbers are obtained via geometry: as lengths of line segments.[26] This requires the use of a congruence relation, say $\cong$, defined over segments. In our case, the line comes as an ordered set, meaning that the line and its segments are *directed*; hence $\alpha\beta \cong \gamma\delta$ also implies: $\alpha > \beta \Leftrightarrow \gamma > \delta$. In Ramsey's notation '=' is also used for congruence; thus he writes: $\alpha\beta = \gamma\delta$. (This agrees with Euclid's terminology and notation, except that in Euclid capital roman letters are use for points, so that, "AB is equal to CD" means that the segment AB is congruent to the segment CD.) Under the identification of $\alpha, \beta, \gamma, \delta, \ldots$ with real numbers, $\alpha\beta = \gamma\delta$, becomes $\alpha - \beta = \gamma - \delta$. Ramsey's idea is to define $\alpha\beta = \gamma\delta$ by means of the following defining condition, where the agent's degree of belief in p is 1/2.

Cong Segments: α *if* p, δ *if* $\neg p$ is equivalent to β *if* p, γ *if* $\neg p$.

[24] "[W]e use Greek letters to represent the different possible totalities of events between which our subject chooses — the ultimate organic unities" (p.176-177).

[25] For example, axiom 3 (p.179) says that the equivalence relation is transitive. Additional properties are implied by the axioms, on the whole.

[26] Or rather, as the ratios of a line segments to some fixed segment chosen as unit.

The underlying heuristics seems to be this: *If* $\alpha, \beta, \gamma, \delta$ are identified with real numbers and *if* (Cong Segments) means that the expected utilities of the two gambles are the same, then an easy computation of expected utilities, for the case in which p is believed to degree 1/2, shows that (Cong Segments) is equivalent to: $\alpha - \beta = \gamma - \delta$. This reasoning presupposes however that the truth (or falsity) of p does not have, by itself, any positive or negative value for the agent. Ramsey calls such propositions *ethically neutral*. The precise, more technical, definition is: an atomic proposition p is ethically neutral "if two possible worlds differing only in regard to the truth of p are always of equal value" (p.177); a non-atomic proposition is ethically neutral if all its atomic components are. Now, if p is ethically neutral, then the agent's having degree of belief 1/2 in p is definable by the condition:

Deg Bel 1/2: For some $\alpha \neq \beta, \alpha$ *if* p, β *if* $\neg p$ is equivalent to β *if* p, α *if*$\neg p$.

Hence, (Cong Segments) can be used to define $\alpha\beta = \gamma\delta$, provided that p is an ethically neutral proposition believed to degree 1/2.

Ramsey's first axiom states that such a proposition exists. Using which, he defines the congruence relation between directed segments and adds further axioms, including the axiom of Archimedes and the continuity axiom, which make it possible to identify the values $\alpha, \beta, \gamma, \ldots$ with real numbers. Applying systematic ambiguity, Ramsey uses the Greek letters also for the corresponding real numbers (and we shall do the same).

Having established this numeric scale of values, Ramsey (p.179-180) proposes the following way of determining a person's degree of belief in the proposition p: Let α, β, γ be such that the following holds.

$p(\alpha, \beta, \gamma)$: α *for certain* is equivalent to β *if* p, γ *if* $\neg p$.[27]

Then the person's degree of belief in p is $(\alpha-\gamma)/(\beta-\gamma)$. Of course, the definition is legitimate iff the last ratio is the same for all triples (α, β, γ) that satisfy $p(\alpha, \beta, \gamma)$. Ramsey observes that this supposition must accompany the definition, that is, we are to treat it as an axiom. A similar axiom is adopted later (p.180) for the definition of conditional degrees of belief, and he refers to them as *axioms of consistency.*

Now the only motivation for adopting the consistency axiom is expediency. The axiom states in a somewhat indirect way that the Greek letters range over a utility scale. Consider the two following claims:

Consist Ax: There is x, such that, for all α, β, γ, $p(\alpha, \beta, \gamma)$ IFF $(\alpha - \gamma)/(\beta - \gamma) = x$.

Utility Scale: There is x, such that, for all α, β, γ, $p(\alpha, \beta, \gamma)$ IFF $\alpha = x \cdot \beta + (1 - x) \cdot \gamma$.

In both claims 'p' is a free variable ranging over propositions, which has to be quantified universally. The second claim states that the value scale established using all the previous axioms is a utility scale—where the number x, which is associated with the proposition p is its subjective probability; (i.e., there is no problem of marginal utility and the acceptance of a bet depends only on the betting odds). Now, by elementary algebra, the two claims

[27] If p is not ethically neutral then the gamble is supposed to be adjusted already, so that β contains the contribution of p and γ — the contribution of $\neg p$.

are equivalent. This means that the consistency axiom is a disguised form of the claim that there is a function that associates with each proposition a degree of belief, such that the value scale over which the Greek letters range is a utility scale.

Ramsey goes on to define conditional probability, using conditional gambles, which comes with its associated consistency axiom. This is followed by a proof that the degrees of belief satisfy the axioms of a finitely additive probability, and some other properties of conditional probability.

To sum up, Ramsey's goal was to show how subjective probabilities can, in principle, be derived from betting behavior (where the stakes are are defined in terms of a suitable utility scale). His excessively strong axioms are motivated largely by this goal.

B Savage's Postulates

We provide a list of Savage's postulates. They are stated using the concepts and notations introduced in Section 0 and Section 2.1 together with the following notions of *conditional preference* and *null events*:

Definition B.1 (Conditional preference)**.** Let E be some event, then, given acts $f, g \in \mathcal{A}$, f is said to be weakly preferred to g *given* E, written $f \succcurlyeq_E g$, if, for *all* pairs of acts $f', g' \in \mathcal{A}$, we have

1. f agrees with f' and g agrees with g' on E,
2. f' agrees with g' on $\overline{E}$, and
3. $f' \succcurlyeq g'$.

That is, $f \succcurlyeq_E g$ if, for all $f', g' \in \mathcal{A}$,

$$\left.\begin{array}{ll} f(s) = f'(s),\ g(s) = g'(s) & \text{if } s \in E \\ f'(s) = g'(s) & \text{if } s \in \overline{E}. \end{array}\right\} \implies f' \succcurlyeq g'. \tag{B.1}$$

In other words, the conditional preference of f over g on the occurrence of event E is defined in terms of all unconditional preferences of f' over g' under the constraints that f' and g' agree, respectively, with f and g on event E and with each other on $\overline{E}$, and that f' unconditionally and weakly preferred to g'.

Definition B.2 (Null events)**.** An event E is said to be a *null* if, for any acts $f, g \in \mathcal{A}$,

$$f \succcurlyeq_E g. \tag{B.2}$$

That is, an event is null if the agent is indifferent between any two acts *given* E. Intuitively speaking, null events are those events that the agent believes that it is improbable that they will obtain.

Savage's Postulates

P1 $\succcurlyeq$ is a weak order (complete preorder).

P2 For any $f, g \in \mathcal{A}$ and for any $E \in \mathcal{B}$, $f \succcurlyeq_E g$ or $g \succcurlyeq_E f$.

P3 For any $a, b \in X$ and for any non-null event $E \in \mathcal{B}$, $\mathfrak{c}_a \succcurlyeq_E \mathfrak{c}_b$ if and only if $a \succcurlyeq b$.

P4 For any $a, b, c, d \in C$ satisfying $a \succcurlyeq b$ and $c \succcurlyeq d$ and for any events $E, F \in \mathcal{B}$, $\mathfrak{c}_a|E + \mathfrak{c}_b|\overline{E} \succcurlyeq \mathfrak{c}_a|F + \mathfrak{c}_b|\overline{F}$ if and only if $\mathfrak{c}_c|E + \mathfrak{c}_d|\overline{E} \succcurlyeq \mathfrak{c}_c|F + \mathfrak{c}_d|\overline{F}$.

P5 For some constant acts $\mathfrak{c}_a, \mathfrak{c}_b \in \mathcal{A}$, $\mathfrak{c}_b \succ \mathfrak{c}_a$.

P6 For any $f, g \in \mathcal{A}$ and for any $a \in C$, if $f \succ g$ then there is a finite partition $\{P_i\}_{i=1}^n$ such that, for all i, $\mathfrak{c}_a|P_i + f|\overline{P_i} \succ g$ and $f \succ \mathfrak{c}_a|P_i + g|\overline{P_i}$.

P7 For any event $E \in \mathcal{B}$, if $f \succcurlyeq_E \mathfrak{c}_{g(s)}$ for all $s \in E$ then $f \succcurlyeq_E g$.

References

de Finetti, B. (1937a). Foresight: Its logical laws, its subjective sources. In Kyburg, H. E. and Smokler, H. E., editors, *Studies in Subjective Probability*, pages 53–118. Robert E. Krieger Publishing Co., Inc., 1980, Huntington, N.Y., 2d edition.

de Finetti, B. (1937b). La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré.*, 7:1–68.

Fishburn, P. C. (1970). *Utility Theory for Decision Making*. Wiley, New York.

Fishburn, P. C. (1981). Subjective expected utility: A review of normative theories. *Theory and Decision*, 13(2):139–199.

Fishburn, P. C. (1982). *The Foundations of Expected Utility*. Reidel Dordrecht.

Gaifman, H. (2004). Reasoning with limited resources and assigning probabilities to arithmetical statements. *Synthese*, 140(1-2):97–119.

Gaifman, H. and Liu, Y. (2015). Context-dependent utilities: A solution to the problem of constant acts in Savage. In van der Hoek, W., Holliday, W. H., and Wang, W. F., editors, *Proceedings of the Fifth International Workshop on Logic, Rationality, and Interaction*, volume LNCS 9394, pages 90–101. Springer-Verlag Berlin Heidelber.

Hacking, I. (1967). Slightly more realistic personal probability. *Philosophy of Science*, pages 311–325.

Jeffrey, R. C. (1965). *The Logic of Decision*. McGraw-Hill, New York.

Jeffrey, R. C. (1983). *The Logic of Decision*. University of Chicago Press, Chicago, 2nd edition.

Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge Studies in Probability, Induction, and Decision Theory. Cambridge University Press, Cambridge, UK.

Koopman, B. O. (1940a). The axioms and algebra of intuitive probability. *Annals of Mathematics*, pages 269–292.

Koopman, B. O. (1940b). The bases of probability. *Bulletin of the American Mathematical Society*, 46(10):763–774.

Koopman, B. O. (1941). Intuitive probabilities and sequences. *Annals of Mathematics*, pages 169–187.

Kraft, C., Pratt, J., and Seidenberg, A. (1959). Intuitive probability on finite sets. *The Annals of Mathematical Statistics*, 30(2):408–419.

Kreps, D. M. (1988). *Notes on the Theory of Choice*. Underground classics in economics. Westview Press, Boulder.

Luce, R. D. and Krantz, D. H. (1971). Conditional expected utility. *Econometrica: Journal of the Econometric Society*, 39(2):253–271.

Pratt, J. W. (1974). Some comments on some axioms for decision making under uncertainty. In Balch, M., McFadden, D., and Wu, S., editors, *Essays on economic behavior under uncertainty*, pages 82–92. American Elsevier Pub. Co.

Ramsey, F. P. (1926). Truth and probability. In Braithwaite, R. B., editor, *The Foundations of Mathematics and other Logical Essays*, number 158–198, chapter VII. London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Company, 1931.

Savage, L. J. (1954). *The Foundations of Statistics*. John Wiley & Sons, Inc.

Savage, L. J. (1972). *The Foundations of Statistics*. Dover Publications, Inc., second revised edition.

Seidenfeld, T. and Schervish, M. (1983). A conflict between finite additivity and avoiding dutch book. *Philosophy of Science*, pages 398–412.

Shafer, G. (1986). Savage revisited. *Statistical Science*, 1(4):pp. 463–485.

von Neumann, J. and Morgenstern, O. (1944). *The Theory of Games and Economic Behavior*. Princeton University Press.