Savage Revisited

Author(s): Glenn Shafer

Source: *Statistical Science* , Nov., 1986, Vol. 1, No. 4 (Nov., 1986), pp. 463-485

Published by: Institute of Mathematical Statistics

Stable URL: https://www.jstor.org/stable/2245794

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve and extend access to *Statistical Science*

This content downloaded from 150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

Statistical Science
1986, Vol. 1, No. 4, 463–501

Savage Revisited

Glenn Shafer

Abstract. Three decades ago L. J. Savage published *The Foundations of Statistics*, in which he argued that it is normative to make choices that maximize subjective expected utility. Savage based his argument on a set of postulates for rational behavior. Empirical research during the past three decades has shown that people often violate these postulates, but it is widely believed that this is irrelevant to Savage's argument.

This article re-examines Savage's argument and concludes that his postulates cannot be so thoroughly insulated from the empirical facts. The argument actually relies heavily on assumptions that have been empirically refuted. Savage's normative interpretation of subjective expected utility must therefore be revised.

The revision suggested here emphasizes the constructive nature of probability and preference. It also emphasizes the constructive nature of small worlds, the frameworks within which probability and utility judgments are made.

According to the constructive understanding, an analysis of a decision problem by subjective expected utility is merely an argument, an argument that compares that decision problem to the decision problem of a gambler in a pure game of chance. This argument by analogy may or may not be cogent. In some cases other arguments are more cogent.

Key words and phrases: Constructive decision theory, normative decision theory, subjective expected utility, subjective probability, sure thing principle.

1. INTRODUCTION

More than three decades have passed since 1954, when L. J. Savage published *The Foundations of Statistics*. The controversy raised by this book and Savage's subsequent writings is now part of the past. Many statisticians now use Savage's idea of personal probability in their practical and theoretical work, and most of the others have made their peace with the idea in one way or another. Thus the time may be ripe for a re-examination of Savage's argument for subjective expected utility.

Savage's argument begins with a set of postulates for preferences among acts. Savage believed that a rational person's preferences should satisfy these postulates, and he showed that these postulates imply that the preferences agree with a ranking by subjective expected utility. He concluded that it is normative to make choices that maximize subjective expected utility. To do otherwise is to violate a canon of rationality.

In the 1950s and 1960s, Savage's understanding of subjective expected utility played an important role in freeing subjective probability judgment from the strictures of an exaggerated frequentist philosophy of probability. Today, however, it no longer plays this progressive role. The need for subjective judgment is now widely understood. Increasingly, the idea that subjective expected utility is uniquely normative plays only a regressive role; it obstructs the development and understanding of alternative tools for subjective judgment of probability and value.

In this article, I shall advocate a revision of Savage's understanding. According to this revision, the analysis of a decision problem by subjective expected utility is merely an argument by analogy. It draws an analogy between that decision problem and the problem of a gambler who must decide how to bet in a pure game of chance. Sometimes such arguments are cogent; sometimes they are not. Sometimes other kinds of arguments provide a better basis for choosing among acts. Thus subjective expected utility is just one of several possible tools for constructing a decision.

1.1 Savage's Normative Interpretation

Savage distinguished between two interpretations for his postulates, an empirical interpretation and a

Glenn Shafer is Professor, School of Business, University of Kansas, Lawrence, Kansas 66045.

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

normative interpretation. According to the empirical interpretation, people's preferences among acts generally obey the postulates and hence agree with a ranking by subjective expected utility. According to the normative interpretation, the postulates are a model of rationality. They describe the preferences of an ideal rational person, an imaginary person whose behavior provides a standard or norm for the behavior of real people. The normative interpretation does not assert that the preferences of real people obey the postulates; it asserts only that they should.

Savage was sympathetic to the empirical interpretation; he thought people's preferences usually come close to obeying the postulates (see Friedman and Savage, 1952; Savage, 1952, page 29; Savage, 1954, page 20; Savage, 1971.) But his primary emphasis, especially in *Foundations*, was on the normative interpretation.

Savage's distinction between empirical and normative interpretations was immensely influential. He was not quite the first to make such a distinction; Jacob Marschak discussed the "descriptive" and "recommendatory" aspects of expected utility in 1950. But Savage's forceful advocacy of the normative interpretation made the distinction widely appreciated. There was scarcely a hint of the distinction in the three editions of von Neumann and Morgenstern's *Theory of Games and Economic Behavior* (1944, 1947, 1953), yet it is difficult to find a discussion of expected utility written after 1954 that does not acknowledge the importance of the distinction.

The normative interpretation has become steadily more important during the past three decades as psychologists have shown in more and more detail that the empirical interpretation is false. It has also become purer. Our careful look back at Savage's words will show us that he scarcely hid the dependence of his argument on what he took to be empirical facts about people's preferences. But today's Bayesian statisticians often contend that empirical facts are completely irrelevant to the normative interpretation. People should obey Savage's postulates, and what they actually do has no relevance to this imperative (Lindley, 1974).

I shall argue that this is wrong. The normative interpretation cannot be so thoroughly insulated from empirical fact. Savage's argument for the normativeness of his postulates cannot be made without assumptions that have empirical content, and what we have learned in the past three decades refutes these assumptions just as clearly as it refutes the forthright empirical interpretation of the postulates. The sensible way to respond to what we have learned is to make the normative interpretation explicitly and thoroughly constructive. This means repudiating the claim that subjective expected utility provides a uniquely normative way of constructing decisions. It may also mean abandoning the word *normative* in favor of *constructive* and other less contentious terms.

1.2 The Existence and Construction of Preferences

Just what are the assumptions with empirical content that underlie Savage's argument for the normativeness of subjective expected utility?

There are at least two. First, the assumption that a person always has well-defined preferences in those settings where the postulates are applied. Second, the assumption that a setting can be found that permits a disentanglement of belief and value.

In order to understand the role of the first assumption in Savage's argument, consider his treatment of the idea that preferences should be transitive. Transitivity seems inherent in the idea of preference; I would be using words oddly if I were to say that I prefer f to g, g to h, and h to f. It would be unreasonable, *prima facie*, for me to insist on using words so oddly. In this sense, transitivity is normative. But Savage went a step further; he declared that it is always normative to have transitive preferences among f, g, and h. This further step is justified if we make the assumption that a person does have preferences among f, g, and h, for then we are merely saying that these preferences, that the person does have, should be transitive. But the further step is not justified if the person does not necessarily already have preferences among f, g, and h. For in this case, we are saying that the person should construct such preferences regardless of how difficult this might be, regardless of how useful it might be, and regardless of what other ways the person might have of spending his or her time.

Psychologists have found that people are usually willing to comply with requests that they choose among options. So how can I claim that the assumption that a person always has preferences is counter to the facts?

Let us reflect on what we need in order to say that an object has a certain property. We need a method or methods of measurement, and we need an empirical invariance in the results of applying these methods. We are entitled to say that a table has a certain length because we have methods of measuring this length and because we get about the same answer from different methods and on different occasions.

In the case of people's preferences, we have methods of measurement. There are questions we can ask. But do we find the requisite empirical invariance? In general, we do not. Trite as this may be, it is the most fundamental result of three decades of empirical investigation. The preferences people express are unsta-

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

ble (Fischhoff, Slovic, and Lichtenstein, 1980). They depend on the questions asked. A person's choice between f and g may depend on whether the conversation includes consideration of h or k (Tversky, 1972). It may also depend on substantively irrelevant aspects of the descriptions of the options, even when these options are treated evenhandedly (Tversky and Kahneman, 1986).

Savage's second assumption with empirical content, the disentanglement of belief and value, is more subtle. The domains of belief and value can be conceptually disentangled for the gambler in a pure game of chance. The gambler has beliefs about the outcome of the game, and he puts values on the different amounts of money he can win from the game. These two domains are initially separate; they are linked only by the gambler's choice of bets. When we analyze a decision problem by subjective expected utility, we are either assuming or deciding on a similar disentanglement. The assumption has empirical content, and the decision may or may not be one that we want to make. I shall argue that the empirical facts do not support the assumption.

The assumption that a person always has well-defined preferences is explicit in Savage's first postulate. We will study this postulate in detail in Section 3. The assumption that a person can frame a decision problem so as to disentangle judgments of value from judgments of probability underlies the second, third, and fourth postulates. We will study these postulates in Section 4.

1.3 The Car Radio

Before we plunge into the technicalities of Savage's postulates, let us think about a more general ingredient of his argument: the idea that preferences can be treated as errors. Savage believed that the normative force of his postulates is such that if a person discovers that his or her preferences violate the postulates, he or she will think of the violation as an error and will change some or all of the preferences so as to correct this error. Here is a simple story he used to illustrate this idea of treating a preference as an error to be corrected:

> ... A man buying a car for $2134.56 is tempted to order it with a radio installed, which will bring the total price to $2228.41, feeling that the difference is trifling. But when he reflects that, if he already had the car, he certainly would not spend $93.85 for a radio for it, he realizes that he has made an error.
>
> *Foundations*, page 103

How does the constructive attitude that I am advocating apply to this story?

From the constructive viewpoint, this story is simply an example of the empirical fact that preferences are not invariant with respect to the method of measurement. The man has asked himself in two different ways what value he puts on a car radio, and he has received two different answers. This means that he does not really have a well-defined preference between the car radio and $93.85. His task is to construct such a preference.

Savage's way of resolving the story suggests that the second question the man asks himself is the right one. When he asks himself directly whether the radio is worth $93.85, he finds that it is not, and this tells him that his initial inclination to pay $93.85 more to have the radio in the car was an error.

But it is equally open to the man to decide that he likes the first question best. He may decide that it is in the context of buying a car that he best faces up to the value he is willing to place on the amenities in the car, and that the discomfort he would feel in paying $93.85 just for a radio causes him to unreasonably undervalue the amenity provided by the radio when he considers it in isolation. In this case, he might call his feeling that he would not pay $93.85 for the radio the error.

The word error is inappropriate here. It suggests that the man's true preference is well-defined before he deliberates and that he just needs to ask himself the right question in order to find out this true preference; other questions may produce errors. From our constructive viewpoint, we see quite a different picture. The man does not really have a true preference, and he is looking to various arguments (including those provided by the salesman) in an effort to construct one.

These considerations involve, perhaps, only a shallow challenge to Savage's viewpoint. I am merely criticizing his casual use of the word error. So let us move to higher ground and ask what is normative in this man's situation.

There is an obvious response. The man has given inconsistent answers to the two questions, and it is normative for him to resolve this inconsistency. It is normative for him to have a clear preference between the radio and $93.85, so that he can henceforth answer the two questions consistently.

There is a sense in which this is correct. The man has to decide whether to pay extra for the radio or not. But it is important to recognize that this is a contingent necessity. It results not from logic but from the fact that the car is available with the radio installed and the salesman has asked him whether he wants it that way. Were it unavailable, the man might have something better to do than to construct a preference between the radio and $93.85.

In a final attempt to find a role for the word normative in this story, one might suggest that if the man must decide whether to pay extra for the radio, then

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

it is normative for him to ask both questions and to reflect on the inconsistency of the answers before taking action. It is normative to look at a decision from all points of view, one might argue, precisely because one's true preference is not well-defined. If we recognize the fuzziness of our preferences and ask ourselves about our preferences in many different ways, then we will be likely to make better decisions than if we act on our answer to the first question we ask ourselves.

Even here, however, *normative* is too strong. It is normative, perhaps, to deliberate carefully. But this says very little. It is never possible to look at a decision from all points of view. And whether and in what sense a decision will be improved by consideration of any particular additional way of asking oneself about one's preference may be an open question. Having asked himself whether the radio is worth the extra money, the man may or may not improve his deliberation by asking himself whether he would pay that much for the radio if he already had the car.

Our discussion of Savage's postulates will involve issues similar to those raised by this simple story.

1.4 Outline

In Section 2, I review the mathematical formulation of Savage's theory. I also discuss the significance of Savage's representation theorem and the ways in which Savage's perspective on subjective expected utility differed from a constructive perspective.

Then, in Sections 3 and 4, I look in detail at Savage's postulates, at the criticisms other authors have made of them, and at their constructive significance. Section 3 is devoted to the first postulate, the requirement that acts be completely ranked in preference. This is the simplest and most important of the postulates. Section 4 is devoted to the second, third, and fourth postulates, which formalize the idea that belief and value can be disentangled in all decision problems as they can be in a gambler's decision problem.

In Section 5, I study Savage's problem of small worlds, contrasting his treatment of this problem with a more constructive treatment. A small world consists of the possible states of the world and the possible consequences that a person considers when he or she analyzes a decision problem. States of the world and consequences must necessarily be described at some fixed and therefore limited level of detail; hence the adjective small. A person can always consider a more refined small world, one with more detailed and hence more numerous descriptions of the possibilities. The problem of small worlds is that an analysis using one small world may fail to agree with an analysis using a more refined small world. From the constructive viewpoint, this is merely one aspect of the lack of invariance of preference; the preferences we construct may depend on which questions we ask ourselves, and hence the selection of questions is an essential part of the construction. Since he implicitly assumed the preexistence of well-defined preferences, Savage found the problem of small worlds more mysterious than this. In fact, Savage's treatment of the problem can serve as a demonstration of how far from a constructive perspective he was.

2. SAVAGE'S THEORY

This section reviews the mathematical formulation of Savage's theory. I review what Savage meant by a small world. I state Savage's seven postulates in a form slightly different from the form in which he gave them in *Foundations*. Then I discuss the representation theorem that Savage deduced from these postulates, its significance from Savage's point of view, and its significance from a more constructive point of view.

2.1 Small Worlds

Suppose I must choose an act from a set F_0 of possible acts, and suppose the consequences of these acts are uncertain. How might I choose?

Savage suggested that I begin by spelling out the possibilities for those present and future aspects of my situation which will be unaffected by my choice of an act but which, together with this choice, will determine the personal consequences that I want to take into account.

Let S denote the set of these possibilities. More concretely, suppose S is a set of written descriptions. Each element s of S describes one way the unknowns in my situation might turn out, in enough detail to determine the relevant consequences of each act. Let us also suppose that the elements of S are mutually exclusive and collectively exhaustive. One and only one of these elements describes my situation correctly. We may call each element of S a possible state of the world.

Let C denote the set of the consequences. Again, we may be more concrete by supposing that C is a set of written descriptions; each element c of C describes one way the personal consequences of my choice of an act might turn out. Let us suppose that the elements of C are mutually exclusive and collectively exhaustive; one and only one of these elements describes what will actually happen to me. For each element s in S and each act f in F_0, let $f(s)$ denote the element of C that correctly describes the personal consequences of the act f if s correctly describes my situation. As the notation indicates, each act in F_0 determines a mapping from S to C.

Savage called the pair (S, c) a *small world.*

On pages 13 to 15 of *Foundations*, Savage formulates a small world for a man who must decide whether to

This content downloaded from 150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

TABLE 1
Savage's small world

Act	State	
	Good	Rotten
Break into bowl	Six-egg omelet	No omelet and five good eggs destroyed
Break into saucer	Six-egg omelet and a saucer to wash	Five-egg omelet and a saucer to wash
Throw away	Five-egg omelet and one good egg destroyed	Five-egg omelet

break a sixth egg into a bowl of five eggs before making an omelet. This is the only small world that Savage completely spelled out in *Foundations*, and it will serve to illustrate some points that we will encounter later.

The man is considering three possible acts:

$$F_0 = \begin{Bmatrix} \text{break the egg into the bowl} \\ \text{break the egg into a saucer} \\ \text{throw the egg away} \end{Bmatrix}.$$

Savage describes the man's situation in terms of a small world (S, C), where S consists of two states of the world, and C consists of six possible consequences. The states of the world simply specify whether the sixth egg is good:

$$S = \begin{Bmatrix} \text{the sixth egg is good} \\ \text{the sixth egg is rotten} \end{Bmatrix}.$$

The consequences specify how large an omelet the man gets in the end, whether he destroys one or more good eggs, and whether he has an extra saucer to wash. Table 1, taken from page 14 of *Foundations*, spells out how the three acts in F_0 map S to C. The act "break the egg into the bowl," for example, maps "the sixth egg is good" to "six-egg omelet" and maps "the sixth egg is rotten" to "no omelet, and five good eggs destroyed."

Savage used this example to illustrate the idea that a person's choice between the acts in F_0 might depend only on which of the consequences in C may befall him. Indeed it might, but do we have any right to demand this? If the man dislikes throwing eggs away without knowing they are rotten, and if he claims the dislike attaches to the act in itself, not just to the misfortune that results if the eggs are not rotten, do we have reason to fault him? We will return to such questions in Section 4.

2.2 The Postulates

Savage's postulates can be stated in a number of equivalent ways. The statement given here is strongly influenced by Fishburn (1981, pages 160 and 161). In order to facilitate the later discussion, I give each postulate a title as well as a number; these titles are mine, not Savage's or Fishburn's.

Consider a small world (S, C) for a set F_0 of possible acts. As we have noted, the relation between F_0 and (S, C) can be expressed by saying that each act f in F_0 determines a mapping from S to C; the mapping that maps the state s to the consequence $f(s)$. If we are content not to distinguish between two acts that have the same consequences, then it is convenient for the abstract theory to identify the act f with this mapping from S to C. The set F_0 then becomes simply a set of mappings. Usually, however, F_0 will not include all mappings from S to C.

Let F denote the set of all mappings from S to C. It is convenient to call all the elements of F *acts*; we may call the elements of F_0 *concrete acts*, and we may call the elements of F that are not in F_0 *imaginary acts*.

Savage's first postulate says that his rational person has ranked in preference all the acts in F, concrete and imaginary:

P1. *The existence of a complete ranking.* All the acts in F are ranked in preference, except that the person may be perfectly indifferent between some acts. More precisely: (i) The binary relation $>$ on F is irreflexive and transitive, where "$f > g$" means that the person prefers f to g. (ii) The binary relation # on F is transitive, where "$f \# g$" means that neither $f > g$ nor $g > f$.

(When we say that $>$ is irreflexive, we mean that $f > g$ and $g > f$ cannot both hold; in particular, $f > f$ cannot hold. When we say that $>$ is transitive, we mean that if $f > g$ and $g > h$, then $f > h$.) The irreflexivity and transitivity of $>$ make precise the idea of a ranking. The transitivity of # makes precise the idea that if neither $f > g$ nor $g > f$, then the person is perfectly indifferent between f and g. Indeed, since $f \# f$ for all f and since $f \# g$ implies $g \# f$, imposing the further condition that # be transitive amounts to requiring that # be an equivalence relation. Thus, the postulate says that F can be divided into equivalence classes, and these equivalence classes can be ranked so that the person prefers acts in equivalence classes higher in the ranking and is indifferent between acts in the same equivalence class.

For each act f in F and each subset A of S, we let f_A denote the restriction of the mapping f to the set A. We call a subset A of S *null* if $f \# g$ whenever f and g are elements of F such that $f_{A^c} = g_{A^c}$, where A^c denotes the complement of A. This condition says that the person's preferences among acts are not influenced by the consequences they have for states in A; we call A null in this case on the presumption that the person's indifference toward A indicates a conviction that the true state of the world is not in A.

Given a subset A of S and two mappings p and q

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

from A to C, let us write $p > q$ if $f > g$ for every pair f and g of mappings in F such that $f_A = p$, $g_A = q$, and $f_{A^c} = g_{A^c}$.

Given a consequence c in C, let $[c]$ denote the act in F that maps all s in S to c. Let us call such an act a *constant act.*

These definitions and conventions allow us to state Savage's remaining postulates as follows:

P2. *The independence postulate.* If $f > g$ and $f_{A^c} = g_{A^c}$, then $f_A > g_A$.

P3. *Value can be purged of belief.* If A is not null, then $[c]_A > [d]_A$ if and only if $[c] > [d]$.

P4. *Belief can be discovered from preference.* Suppose $[c] > [d]$, f is equal to c on A and d on A^c, and g is equal to c on B and d on B^c. Suppose similarly that $[c'] > [d']$, f' is equal to c' on A and d' on A^c, and g' is equal to c' on B and d' on B^c. Then $f > g$ if and only if $f' > g'$.

P5. *The nontriviality condition.* There exists at least one pair of acts in F, say f and g, such that $f > g$.

P6. *The continuity condition.* If $f > g$, then for every element c of C there is a finite partition of S such that f (or g or both) can be changed to equal c on any single element of the partition without changing the preference.

P7. *The dominance condition.* If $f_A > g_A$, then $f_A > [g(s)]_A$ for some s in A, and $[f(s)]_A > g_A$ for some s in A.

These postulates imply that the person's preferences among acts can be represented by subjective expected utility. That is to say, they imply the existence of a probability measure P on S and a real-valued function U on C such that $f > g$ if and only if $E(U(f)) > E(U(g))$, where the expectations are taken with respect to P.

The last three postulates play a relatively technical role in Savage's theory. The nontriviality condition is not needed to prove the representation theorem; it merely assures that the representation is not trivial. The continuity condition is a simplifying or structural assumption; it implies that U is bounded (Fishburn, 1970, page 206). The dominance condition is not needed for the representation theorem in the case of acts that take only finitely many values in C. I will not discuss these three postulates further in this article.

The first four postulates do play significant substantive roles, and I will discuss them in detail, the first postulate in Section 3 and the other four in Section 4.

2.3 The Representation Theorem

Whenever we construct probabilities and utilities and use them to construct a preference ranking for acts, the resulting preferences will satisfy the first four of Savage's postulates. These postulates should therefore be of interest to anyone who takes the constructive view that I set forth in Section 1. They help us understand the limitations of this particular way of constructing a decision. But why should anyone be interested in Savage's representation theorem, which goes in the opposite direction, from preferences to probabilities and utilities?

The representation theorem would be of interest to the constructive view if preferences between acts were a starting point for construction. If, without first constructing probabilities and utilities, a person could state extensive definite preferences satisfying Savage's postulates, then we could use Savage's representation theorem to find probabilities and utilities that would summarize those preferences. Even if the person could only state extensive definite preferences that nearly satisfy the postulates, we might be able to find probabilities and utilities that nearly summarize those preferences, and the person might gain a clearer self-conception by adjusting his preferences so that they fit these probabilities and utilities exactly and hence, incidentally, satisfy the postulates.

Although Savage did not use the word construction in connection with probability and utility, he did think that preferences are the proper starting point for the investigation of a real person's beliefs and values. He thought, for example, that the most effective way to find out about a person's probability for an event is to ask him to choose between bets on the event (Savage, 1971). He thought that a person could, for the most part, express definite preferences between hypothetical acts, and he thought that these preferences would be in close enough accord with his postulates that they could be used to deduce probabilities and utilities (*Foundations*, page 28).

Was Savage right? Do real people, when they have not deliberately constructed probabilities and utilities for a given problem, always have preferences that are sufficiently definite and detailed, and accord well enough with Savage's postulates, that they determine such probabilities and utilities? This is an empirical question, and the empirical studies I have already cited suffice to establish that it must be answered in the negative.

I conclude that Savage's representation theorem is not a constructive tool. In this article I will argue that it is almost always more sensible to construct preferences from judgments of probability and value than to try to work backward from choices between hypothetical acts to judgments of probability and value. Probabilities should be constructed by examining evidence, not by examining one's attitudes toward bets. Utilities are too delicate to be deduced from hypothetical choices; they must be deliberately adopted.

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

3. THE CONSTRUCTIVE NATURE OF PREFERENCE

In this section, we will study Savage's first postulate, which demands that people rank acts in preference. As I have already argued, this demand depends *prima facie* on the claim that they do have fairly well-defined preferences between most pairs of acts. If people do have such preferences, then saying they should have a complete preference ranking amounts only to saying that they should straighten out some inconsistencies and fill in some minor hiatuses, and this may be reasonable. But if they do not have all these preferences, then it is hard to see why constructing them would necessarily be the best way for them to spend their time.

In fact, people generally do not have ready-made preferences. When asked to make choices, they look for arguments on which to base these choices. The ways in which the alternatives are described can suggest arguments and therefore influence these choices. This means that people's choices in response to one query may be inconsistent with their choices in response to another query, but this weak kind of inconsistency is inescapable for rational beings who base their choices on arguments.

Before developing these points in greater detail, let us look more closely at the meaning of the first postulate.

3.1 Indecision and Indifference

The very meaning of preference seems to involve transitivity: if f is preferred to g and g is preferred to h, then f is preferred to h. It is reasonable, therefore, to say that a person who constructs intransitive preferences is being inconsistent. Savage's first postulate demands more, however, than the transitivity of preferences. It also demands transitivity for the binary relation #, which corresponds to lack of preference. Is transitivity involved in the very meaning of lack of preference?

We will be able to understand the significance of transitivity for # more clearly if we formally distinguish between indecision and indifference. Given a person with a transitive and irreflexive preference relation $>$ on F, let us say that the person is *undecided* between f and g if neither $f > g$ nor $g > f$. And let us say that he is *indifferent* between f and g only if in addition to being undecided between them he is also willing to substitute one for the other in any other preference relation. (More precisely, $f > h$ if and only if $g > h$, and $h > f$ if and only if $h > g$.) With this vocabulary established, the significance of transitivity for # is easily stated: # is transitive if and only if the person is indifferent between every pair of acts between which he is undecided.

The demand that a person should be indifferent whenever he is undecided does not seem very reasonable. The person might be undecided between two acts because he feels he lacks the evidence needed for a wise choice, because he feels the choice depends on more fundamental choices or value judgments not yet made, or simply because he feels the choice is one he does not need to make. Indifference says much more.

For the constructive view, indecision is the starting point. Before we start to work constructing preferences, we may be undecided between all pairs of acts. We may not even have thought of all the possible acts. But this does not mean we are indifferent. As we construct preferences, we eliminate some indecision. In the end we may eliminate all indecision; we may, that is to say, rank all acts in a strict order of preference. Or we may, as the postulate suggests, reduce all indecision to indifference, by establishing a ranking of equivalence classes of acts. But Savage has given us no reason why we should feel compelled to carry our elimination of indecision so far. In general, the practical problem will be to choose one act. Why is it normative to go further and rank all acts?

These points were not overlooked by Savage's early critics. The main points were made quite well by Anscombe (1956), Aumann (1962, 1964), and Wolfowitz (1962). Anscombe and Wolfowitz were primarily concerned with statistical problems. Citing the problem of choosing a statistical model, Anscombe made the point that we sometimes cannot even list all the possible choices that are open to use, let alone rank them. Wolfowitz made the point that in a practical problem of choice, there is a practical need to choose a single act to perform, but no practical need to rank all the other acts. He suggested that the unreasonableness of Savage's demand that a person rank all acts could be illustrated by

> . . . a homely example of the sort which Professor Savage uses frequently and effectively: When a man marries he presumably chooses, from among possible women, that one whom he likes best. Need he necessarily be able also to order the others in order of preference?
>
> Wolfowitz, 1962, page 476

If we were to assume that a man or woman, when thinking about marriage, begins with well-defined preferences between every pair of possible spouses, then it would be reasonable to ask that these preferences be transitive. But there are no grounds for this assumption. And there is also no compelling reason for the person to try to construct such a ranking.

The distinction between indecision and indifference is not as clear as it might be in Savage's own discussion of his first postulate, primarily because he expressed the postulate in terms of the relation "is not preferred to." We say that f is not preferred to g, or $f \leq g$, if and only if $f > g$ does not hold. Savage imposed two conditions on $\leq$: (i) for any pair of acts f and g,

This content downloaded from 150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

at least one of the relations $f \leq g$ or $g \leq f$ holds, and (ii) $\leq$ is transitive. It is obvious that (i) is equivalent to $>$ being irreflexive. It is also true, but not so obvious, that (ii) is equivalent to both $>$ and $\#$ being transitive.

3.2 Where Should We Put Our Effort?

In response to Wolfowitz's point, that it is unnecessary to rank alternatives we are not going to choose, some readers will point out that the exercise of constructing such a ranking may help us better understand values that we do have. Indeed it may. But is there any reason to suppose that it will always do so? And is there any reason to suppose that this exercise is always the best way we can use our time?

Instead of trying to rank in order all the men she dislikes, a woman might better spend her time learning more about the man she favors. Or perhaps she should spend her time exploring her possibilities in terms of a more detailed small world, one that relates her possible choice of a husband to other choices.

In my view, we can never say that it is normative for a person to construct a complete preference ranking of the acts in a given small world, because we can never be certain that this is the best way for the person to spend his or her time. It may be better to spend this time looking for further evidence. It may be better to spend it trying to invent other small worlds that provide more convincing frameworks for probability and value judgment. Or it may be time to put an end to deliberation and get on with one's life.

3.3 The Empirical Claim

Savage acknowledged the possibility of distinguishing between indecision and indifference in the following words:

> There is some temptation to explore the possibilities of analyzing preference among acts as a partial ordering, that is, in effect to replace [the requirement that $f \leq g$ or $g \leq f$] by the very weak proposition $f \leq f$, admitting that some pairs of acts are incomparable. This would seem to give expression to introspective sensations of indecision or vacillation, which we may be reluctant to identify with indifference. My own conjecture is that it would prove a blind alley losing much in power and advancing little, if at all, in realism; but only an enthusiastic exploration could shed real light on the question.
>
> *Foundations*, page 21

This admirably undogmatic statement comes at the end of a passage in which Savage explains that it is the normative rather than the empirical interpretation of his postulates that has direct relevance to his argument. Yet comments about realism and introspective sensations of indecision are clearly comments about empirical facts, not about what is merely normative. We may take this passage as a concession that the normative interpretation has empirical content.

My contention that Savage based his normative interpretation on the assumption that his first postulate has substantial empirical validity is supported by his article on the elicitation of probabilities and expectations (Savage, 1971), where he asserts that a real person is approximately like a *homo economicus*, who does have ready-made preferences among gambles. Moreover, Savage repeatedly said that the way to use his theory is to search for intransitivities and other inconsistencies in one's preferences and then revise these preferences to eliminate the inconsistencies (see, e.g., Savage, 1967, page 309).

3.4 Constant and Other Imaginary Acts

Some scholars who have been sympathetic with Savage's viewpoint and have accepted the idea that a person should have a complete preference ranking for concrete acts have nonetheless balked at the idea that the person should have a complete preference ranking for imaginary acts. They have been especially concerned about constant acts, acts that map all states of nature to a single consequence. Constant acts play a prominent role in the postulates (postulates P3, P4, and P7 all involve constant acts), but in most small worlds, they are imaginary. Not only that, they are often hard to imagine. It is often hard, that is to say, to imagine performing an act that would result in the consequence c no matter what. And it seems unlikely that people will have in hand preferences between acts that they have not even imagined performing (see Fishburn, 1970; Luce and Krantz, 1971; Pratt, 1974; Richter, 1975).

Savage never published a response to this concern, but his private response, as reported by Fishburn (1981), had a constructive flavor. He saw no reason why a person could not think about patterns of consequences corresponding to imaginary acts and formulate preferences between such patterns.

I agree with Savage on this point. In order to construct a preference between one pattern of consequences and another, it is not necessary that a person should have available a concrete act that produces this pattern, or even that the person should be able to imagine such an act. It makes as much sense for a woman to try to decide which of two men she would prefer as a husband in the case where neither is willing as it does in the case where both are willing but she prefers to marry neither. And as long as she is daydreaming, she might as well also compare these men to imaginary constant husbands, husbands whose qualities and contributions to her life are unaffected by her uncertainties about the state of the world.

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

The scholars who raised the problem of constant acts were identifying an important and valid criticism, however, of the empirical content of Savage's first postulate. While it might be plausible that people have fairly well-defined preferences among the acts available to them, at least in cases where these acts have been present to their imagination for some time, it is less plausible that they have formed such preferences among abstract acts that do not correspond to choices they have thought about.

We will gain some further insight into the problem of imaginary acts when we study the refinement of small worlds in Section 5.2.

3.5 The Empirical Evidence

I contend that Savage's first postulate does not have the degree of empirical validity that it would need in order to be normative. What are the facts? Since 1954 we have accumulated an immense amount of empirical evidence about people's preferences (see, for example, Kahneman, Slovic, and Tversky, 1982; Schoemaker, 1982). Does this evidence show that people always have preferences that are sufficiently definite and extensive that it is reasonable to adjust them so they will satisfy the first postulate perfectly? Or does it show instead that people's preferences are often so fragmentary that there may be better uses of the time and effort needed to make them satisfy it?

This empirical evidence is itself subject to interpretation, of course. It is easy to find people that are willing to participate in experiments where they are required to make many choices, and at first it seems harmless to say that these choices really are their preferences at the time they are announced. This might lead us to agree that people have very extensive preferences. When we then find that these preferences are intransitive and even flatly inconsistent, we are tempted to conclude that it is indeed normative to fix them up so they will be consistent and transitive. But as I pointed out in Section 1.2, the preferences a person expresses often lack the invariance needed to establish them as properties of the person. When we see the extent to which an experimenter influences choices by the way in which he describes alternatives, we realize that the preferences expressed may be more a property of the experiment than a property of the person expressing them.

Let us look at this issue more closely, considering first the claim that people have inconsistent preferences, and then the claim that they have intransitive preferences.

Inconsistent Preferences. Consider the following experiment reported by Tversky and Kahneman (1986).

In the first part of the experiment, participants were asked to choose between two lotteries, A and B. In both lotteries, one randomly draws a marble from a box and wins or loses a sum of money which depends on the color drawn. The percentages of marbles of the different colors and the corresponding gains and losses are given in Table 2. All the participants in the experiment chose lottery B, presumably because they noticed that it gives a better outcome no matter what ball is drawn.

In the second part of the experiment, participants were asked to choose between lotteries C and D given in Table 3. The probability distribution of outcomes is the same for C as for A, and the same for D as for B. So from an abstract point of view, the choice between C and D is the same as the choice between A and B. We can say that B is better than A because the probability distribution of outcomes for B stochastically dominates that for A, and D is better than C for exactly the same reason. But the stochastic dominance is not so easy to see when one is comparing C and D as it is when one is comparing A and B. A majority of the participants in the experiment apparently failed to see it, because they chose C over D.

As this experiment demonstrates, the preferences people express between two probability distributions of gains depend on how the distributions are described. We can express this, if we wish, by saying that people have inconsistent preferences. But it is fairer to say that they do not have any fixed preferences at all. They do not have ready-made answers to the questions asked. Asked to make a choice, they look for arguments. Stochastic dominance is a very convincing argument, if you see it. If you do not see it, then you look for other arguments.

Another remarkable experiment is reported by Tversky and Kahneman (1981). In this experiment, people are told that the United States is preparing for the outbreak of an unusual Asian disease, which is

TABLE 2
A choice between lotteries

Lottery	White	Red	Green	Blue	Yellow
A	90%	6%	1%	1%	2%
	\$0	Win \$45	Win \$30	Lose \$15	Lose \$15
B	90%	6%	1%	1%	2%
	\$0	Win \$45	Win \$45	Lose \$10	Lose \$15

TABLE 3
Another choice between lotteries

Lottery	White	Red	Green	Yellow
C	90%	6%	1%	3%
	\$0	Win \$45	Win \$30	Lose \$15
D	90%	7%	1%	2%
	\$0	Win \$45	Lose \$10	Lose \$15

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

expected to kill 600 people in the absence of any preventive program, and they are asked to choose between two alternative preventive programs. In one case, the possible consequences of the two programs are described as follows:

If Program A is adopted, 200 people will be saved.
If Program B is adopted, there is ⅓ probability that 600 people will be saved, and ⅔ probability that no people will be saved.

In the other case, they are described as follows:

If Program A is adopted, 400 people will die.
If Program B is adopted, there is ⅓ probability that nobody will die, and ⅔ probability that 600 people will die.

The two sets of descriptions are equivalent; 200 people being saved is the same as 400 dying. People choose differently, however, depending on which description is used. In the case of the first description, a large majority of people in the experiment chose Program A, while in the case of the second description a large majority chose Program B. Apparently the first description encourages people to argue in favor of the program that will at least be sure to save some of the people, while the second description encourages them to argue in favor of the program that may result in no deaths at all. Similar results, indicating risk aversion when problems are framed in terms of gains and risk taking when problems are framed in terms of losses, have been obtained when the gains or losses are modest amounts of money rather than lives.

Again, it is possible to say that people are inconsistent because their choice depends on the description of the problem, and depends in particular on the experimenter's choice of a reference point. But it is more helpful to say that the two ways of describing the public health problem suggest different arguments. This is more helpful because it encourages us to weigh the two arguments against each other and to look for other arguments that might help us choose which program to adopt.

Tversky, Kahneman, and others have used these and other experiments to investigate in detail the kinds of arguments that people do use when they make choices. This work is important and relevant to a constructive theory of decision. Here I am making only the elementary point that it is misleading to summarize these experiments by saying that people are inconsistent.

Intransitive Preferences. The study of intransitive preferences goes back at least to Condorcet (1743–1794), who pointed out that a circular pattern of preferences can result from majority voting. Suppose, indeed, that Tom, Dick, and Harry want to decide together among three alternatives A, B, and C. They each rank the alternatives; Tom ranks them ABC (he likes A best and B second best), Dick ranks them BCA, and Harry ranks them CAB. If they vote on each pair, then A will beat B, B will beat C, and C will beat A.

One might expect similar intransitive sets of preferences to be expressed by a single individual who scores his alternatives on several dimensions and chooses between any pair of alternatives by counting the number of dimensions that favor each element of the pair. Tversky (1969), building on a suggestion by May (1954), devised an experiment in which people do consistently produce such intransitivities.

In fact, Tversky (1969, page 32) demonstrates intransitivities with alternatives that differ on only two dimensions. Tversky considers a situation where we are asked to choose between candidates for a job on the basis of their IQ scores and their experience. Suppose we prefer to choose the more intelligent candidate, but we will choose the more experienced candidate if the difference in their IQ scores is negligible. Let d denote the largest difference in IQ scores we consider negligible, and suppose candidates A, B, and C have the IQ scores and experience shown in Table 4. Then we will choose A over B, B over C, and C over A.

Transitivity is so essential to the idea of preference that it does seem reasonable to say that we should reconsider our decision rule. Perhaps instead of regarding d as a negligible difference in IQ scores we should avoid intransitivities by choosing between candidates on the basis of some weighted average of IQ and experience.

If we take a thoroughly constructive view of preference and decision, however, it is important to ask just how widely the decision rule is to be used—i.e., just what preferences are to be constructed. If we want to choose one or more candidates from a pool of three or more, or if we want to repeatedly choose between pairs of candidates, then we may feel that fairness demands a rule that is transitive, even if somewhat arbitrary. But if we face only a single isolated choice, say a choice between candidate A and candidate B, then it may be a waste of time to search for a rule that would seem fair in a wider context.

Here, as always, we must weigh arguments. Given two particular candidates for our job, we may be convinced by the argument that the difference in their IQ scores is negligible. And we may not feel that we

TABLE 4
A choice among three candidates

Candidate	IQ	Experience
		yr
A	100	3
B	100 + d	2
C	100 + 2d	1

This content downloaded from 150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

have enough evidence to construct a convincing argument for a decision rule that uses a particular weighted average of IQ and experience.

"But," the reader may insist, "doesn't it bother you that you are using a rule that produces intransitivities when it is more widely applied?" I must respond that I have enough to worry about as I try to find adequate evidence or good arguments for my particular problem. If I allow myself to be bothered whenever my evidence is inadequate for the solution of a wider problem, then I will always be very bothered. When you call my argument for choosing candidate B over candidate A a rule and choose other situations in which to apply this rule, you are choosing one out of many possible wider contexts in which my argument might be made. This is not reasonable. There are always many wider contexts in which a particular argument might be made, and it is unreasonable that the argument should be convincing in all of them.

The point I am making here is simple: regarding the difference in IQ as negligible may be about the best we can do. I have made the point at length in order to demonstrate how well it can be made when we insist on talking about evidence, argument, and the construction of preference. Matters become much more confused when we try to make the same point using a vocabulary based on the fiction that we already have preferences and that we are just finding out what they are.

4. THE CONSTRUCTIVE NATURE OF SMALL WORLDS

In the small world of the gambler, value is disentangled from probability and belief. The gambler values the amount of money he wins. He has beliefs about the outcome of the game. The two are initially quite distinct; they become connected only when he chooses a gamble. This means that the first step in constructing an argument based on subjective expected utility is to distinguish sharply the consequences on which we want to place value from the questions of fact about which we have evidence. We must construct sets C and S such that we can put utilities on the consequences in C without regard to our evidence about S, and such that we can put probabilities on the states in S without regard to our feelings about C.

According to the constructive view, we may or may not succeed in distinguishing so sharply between a domain of value and a domain of belief. If we do not succeed, then we will have no subjective expected utility argument. We will have to look for other arguments on which to base our decision. According to Savage's normative view, on the other hand, this disentanglement of value and belief is essential to rational decision.

The assumption that value and belief can be disentangled underlies Savage's second, third, and fourth postulates. In this section I contend that Savage made no real case for this assumption. He simply took it for granted.

The second postulate, the independence postulate, has been the most controversial of Savage's postulates. Both its descriptive and normative status have been put in doubt by well-known examples devised by Allais and Ellsberg. I will review these examples and place myself on the side of those who do not find the postualte compelling.

The third and fourth postulates have not received so much attention. They are sometimes said to be uncontroversial. But from a constructive viewpoint, they are more important than the independence postulate, because they express more clearly the assumption that one's small world disentangles value from belief. In order to emphasize this point, I will discuss the third and fourth postulates first, before turning to the independence postulate.

4.1 Can Value Be Purged of Belief?

The third postulate says that if A is not null, then $[c]_A > [d]_A$ if and only if $[c] > [d]$. Recall that $[c]_A > [d]_A$ means that $f > g$ whenever f and g are acts that agree on A^c but satisfy $f(s) = c$ and $g(s) = d$ for s in A; intuitively, this seems to mean that the person prefers the consequence c to the consequence d when his or her choice is limited to the event or situation A. Thus, the postulate says that if the person prefers c to d in general, then he or she prefers it in every situation A. Specializing to the case where A consists of a single state of the world, say $A = \{s\}$, we can say that the person prefers c to d in every state of the world s. Which state of the world is true is irrelevant to the preference.

This postulate clearly expresses one aspect of the disentanglement of value from belief. It says that the question about which we have beliefs (which element of S is the true state of the world?) is irrelevant to our preferences.

The fact that this postulate may fail to hold is brought out by the following example, which Savage gave on page 25 of *Foundations*:

> Before going on a picnic with friends, a person decides to buy a bathing suit or a tennis racket, not having at the moment enough money for both. If we call possession of the tennis racket and possession of the bathing suit consequences, then we must say that the consequences of his decision will be independent of where the picnic is actually held. If the person prefers the bathing suit, this decision would presumably be reversed, if he learned that the picnic were not going to be held near water.

This content downloaded from 150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

Apparently the person prefers the bathing suit to the tennis racket only because he considers it probable that the picnic will be held near water. It seems reasonable that he should reverse his preference when he learns that the facts are otherwise. But this reasonable reversal violates the third postulate. Take A to be the event that the picnic is not going to be held near water, c to be possession of the bathing suit, and d to be possession of the tennis racket. The person's preference for the bathing suit over the tennis racket is indicated by the relation $[c] > [d]$. His preference for the tennis racket when he knows that the true state of the small world is in A is indicated by the relation $[d]_A > [c]_A$.

Savage defended the postulate against this apparent counterexample as follows (again page 25 of *Foundations*):

> ... under the interpretation of "act" and "consequence" I am trying to formulate, this is not the correct analysis of the situation. The possession of the tennis racket and the bathing suit are to be regarded as acts, not consequences. (It would be equivalent and more in accordance with ordinary discourse to say that the coming into possession, or the buying, of them are acts.) The consequences relevant to the decision are such as these: a refreshing swim with friends, sitting on a shadeless beach twiddling a brand new tennis racket while one's friends swim, etc. It seems clear that, if this analysis is carried to its limit, the question at issue [whether $[d]_A > [c]_A$ and $[c] > [d]$ should be allowed] must be answered in the negative ...

The suggestion seems to be that we can always resolve the problem by considering more fundamental consequences. By describing the consequences in a more refined way, we can make their valuation independent of which element of S is true.

The difficulty with this suggestion is that the refinement of C may force a refinement of S. This is because the states of the world in S must be detailed enough to determine which element of C will be achieved by each of our concrete acts. Savage suggests that we take C to consist of descriptions such as "refreshing swim with friends" instead of descriptions such as "possession of bathing suit." But if we want each element of S to determine whether the consequence "refreshing swim with friends" is achieved by the purchase of a bathing suit, we may need to refine S so that its elements say not only whether the picnic will be held near water but also whether the temperature is warm enough for a refreshing swim, which friends come, and so on. And now, since S is more refined, we may face anew the problem of making our preferences among the elements of C independent of which element of S is true. Perhaps the swim will be more refreshing with some friends than others. We face a potential infinite regress, an endless sequence of alternative refinements of C and S.

Another way of putting the matter is to say that we have no reason to suppose that for a given set F_0 of concrete acts we will be able to find S and C such that both (1) each state s in S determines which consequence in C will result from each f in F_0, and (2) the value we want to place on each c in C will not depend on which element of S is the true state. These two desiderata push in opposite directions. The first desiderata pushes us to limit the detail in C or increase the detail in S, while the second pushes us to increase the detail in C or limit the detail in S. There is no *a priori* reason to expect that we can find a compromise that will satisfy both desiderata.

It seems clear, Savage says, that probability and value will finally be disentangled when the "analysis is carried to its limit." This is both lame and vague. In truth, it is not clear what carrying the analysis to its limit would mean, let alone what would happen there. Presumably, carrying the analysis to its limit means looking at ever more refined small worlds, until one arrives at a "grand world," a pair (S, C) so detailed that it takes everything into account. Yet it is hard to make sense of the idea of a grand world.

In Section 5, I will examine Savage's own struggle with the idea of a grand world on pages 82–91 of *Foundations*. Let me remark here that one aspect of the problem is the difficulty in sustaining a distinction between consequences and states of the world as we look at the world in more and more detail. Consequences are states of the person, as opposed to states of the world (*Foundations*, page 14). For some problems, at some levels of detail, I can describe states of my person C and states of the world S in such a way that I care about which state in C happens to me but I do not care about which state in S happens to the world. But when I try to think about very detailed states of the world, states that specify the fate of my own hopes and loved ones, it begins to sound bizarrely hedonistic for me to say that I care not about which of these states happens to the world but only about the consequences for me.

4.2 Can Belief Be Discovered from Preference?

The fourth postulate carries the idea underlying the third postulate a step further. If the relation $[c] > [d]$ does mean that the person values c over d without regard to which element of S is true, then by comparing this absolute preference to the person's other preferences among acts, we can learn about his beliefs about which element of S is true.

Suppose, indeed, that $[c] > [d]$, f is equal to c on A and to d on A^c, and g is equal to c on B and to d on

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

B^c. And suppose that $f > g$. If we assume that value in our small world has been purged of belief—if, that is to say, the preference for c over d is independent of whether the true state of nature is in A and of whether it is in B—then the only available explanation for the preference $f > g$ is that the person considers A more probable than B.

In order for this to work, however, the preference $f > g$ must be unchanged when c and d are replaced by any other pair of consequences c' and d' such that $[c'] > [d']$. As Savage put it, "on which of two events the person will choose to stake a given prize does not depend on the prize itself" (*Foundations*, page 31). The fourth postulate posits that this is the case.

It is easy to create examples, analogous to the example of the tennis racket and bathing suit, in which the fourth postulate does not hold. There is no need to dwell on such examples here. It is worthwhile, though, to reiterate that this postulate derives its force from the assumption that the small world disentangles belief from value. The postulate does not have any normative appeal—it is not even comprehensible—until this assumption is made.

4.3 The Independence Postulate

Consider an act f and a subset A of the set of states of a small world. Imagine changing the consequences that f would have if the true state of the small world were in A—i.e., imagine changing the values $f(s)$ for s in A. This changes f to a different act, say g. The act g differs from f on A but agrees with f on A^c. The change from f to g may be a change for the worse—i.e., we may have $f > g$. Savage's second postulate, the independence postulate, says that whether it is a change for the worse is independent of the consequences that f has under the other states, those in A^c. In other words, if f' is any act that agrees with f on A, and we change f' in the same way that we changed f, thus obtaining an act g' that agrees with g on A but with f' on A^c, then $f > g$ if and only if $f' > g'$.

Here are some other ways of expressing the independence postulate: (1) *More verbally*, if two acts agree on A^c, then the choice between them should depend only on how they differ on A; it should not depend on how they agree on A^c. (2) *More succinctly*, if $f > g$, $f'_A = f_A$, $g'_A = g_A$, $f_{A^c} = g_{A^c}$, and $f'_{A^c} = g'_{A^c}$, then $f' > g'$. (3) *Yet more succinctly, as in Section 3.1 above*, if $f > g$ and $f_{A^c} = g_{A^c}$, then $f_A > g_A$.

In Section 4.3.1, I present two examples that have inspired much of the discussion of Savage's independence postulate. One of these was devised by Maurice Allais, the other by Daniel Ellsberg. Both are counterexamples to the empirical validity of the independence postulate, inasmuch as most people (including Savage himself; see *Foundations*, page 103; Allais, 1979, page 533; Ellsberg, 1961, page 656) express preferences that violate the postulate when they first encounter the examples.

After reviewing the counterexamples, I discuss a number of arguments that have been offered for the independence postulate and against the counterexamples. In Section 4.3.2, I discuss the assertion made by Oskar Morgenstern and others that reasonable people will correct their preferences to conform to the postulate when divergences are pointed out to them. In Section 4.3.3, I discuss the "sure thing principle," the intuitive principle on which Savage based his case for the independence postulate. In Sections 4.3.4 and 4.3.5, I discuss Howard Raiffa's arguments. Finally, in Section 4.3.6, I question Paul Samuelson's contrast between commodities and states of a small world. I contend that goals can tie together states of the world just as they tie together commodities.

4.3.1 The Counterexamples

The two examples presented here are from Allais (1953) and Ellsberg (1961), respectively. It is generally agreed that Allais's is the more important of the two. Ellsberg's example turns on subtle issues about the knowledge of chances, whereas Allais's, although it is usually presented in a chance setting, does not really depend on the idea of chance. Furthermore, the argument for violating the independence postulate is stronger in Allais's example, because the goal thereby attained is more attractive.

Allais's Example. Consider a small world that has three states and has monetary prizes as consequences. The states are s, t, and u, and the prizes are \$0, \$500,000, and \$2,500,000. Consider the acts f, g, f', and g' given in Table 5. If we set $A = \{s, t\}$, then these acts satisfy $f'_A = f_A$, $g'_A = g_A$, $f_{A^c} = g_{A^c}$, and $f'_{A^c} = g'_{A^c}$. The independence postulate therefore forbids us to prefer f to g and g' to f'.

Suppose, however, that we think the true state of the small world is probably u and almost certainly either t or u. (In the version of the example reported in *Foundations* (pages 101–103), we assign probability .01 to s, probability .10 to t, and probabiliity .89 to u.) In this situation, most people violate the postulate by preferring f to g and g' to f'. When comparing f to g, they reason that they can gain \$500,000 for sure by choosing f, and they do not want to risk this very attractive sure thing by gambling for more. But when

TABLE 5
Allais's example

	s	t	u
f	\$500,000	\$500,000	\$500,000
g	\$0	\$2,500,000	\$500,000
f'	\$500,000	\$500,000	\$0
g'	\$0	\$2,500,000	\$0

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

comparing f' to g', they realize that they are likely to get nothing at all, and feeling that they have less to lose, they are more willing to gamble for the larger prize.

One way of putting this is to say that there is a strong argument for choosing f over g which is not available when we compare f' and g'. Another way of putting it is to say that the choice between f and g gives us an opportunity to adopt and attain a goal: the acquisition of $500,000.

It is worth emphasizing that the force of the example does not depend on assigning probabilities to s, t, and u. It is quite enough to say that there is strong evidence for u and even stronger evidence against s.

Allais's example is sometimes presented simply in terms of payoffs and probabilities as in Table 6. When it is presented in this way, we cannot say that the preferences $f > g$ and $g' > f'$ violate the independence postulate, since we are not working in a small world in Savage's sense. It is impossible, however, to assign utilities to the dollar payoffs so that f will exceed g and g' will exceed f' in expected utility.

Readers of Savage's account of the example (*Foundations*, pages 101–103) sometimes gain the impression that Allais originally presented it simply in terms of payoffs and probabilities and that it was Savage who recast it in terms that made the preferences $f > g$ and $g' > f'$ directly contradict the independence postulate. This is not correct, however. In Allais (1953), the example is explicitly presented as a counterexample to the independence postulate (see Allais and Hagen, 1979, pages 88–90 and note 240 on page 586).

Ellsberg's Example. Consider another small world with three states, but with a more modest prize: $100. The acts are shown in Table 7. Here, as in Allais's example, the independence postulate forbids us to prefer f to g and g' to f'. It also forbids us to prefer g to f and f' to g'.

In this case, the forbidden preferences are produced by assuming partial knowledge of objective chances for the state of the small world. Suppose we know that this state is determined by drawing a ball from an urn containing 90 balls. We know that exactly 30 of these balls are labeled s. We know that each of the other 60 is labeled either t or u, but we have no evidence about the proportion.

We know that f offers a $2/3$ chance at the $100 prize. We do not know exactly what chance g offers; we know only that it is between $1/3$ and 1. When offered a choice between f and g, some people say they are completely indifferent. They reason that since there is no reason to think that there are more balls labeled t than u or more labeled u than t, the subjective probability of getting the prize from g is $2/3$, the same as the probability of getting it from f. But most people are not indifferent. Many prefer f to g, because f offers more security; these are the pessimists. Others, the optimists, prefer g to f because g offers the possibility of a greater chance at the $100.

Most people also see a difference between g', which offers a $1/3$ chance at the $100, and f', which offers an unknown chance between 0 and $2/3$. The pessimists, those who chose f over g, choose g' over f'. The optimists, those who chose g over f, choose f' over g'. Both the pessimists and the optimists violate the independence postulate.

The argument for violating the independence postulate is not as strong in this example as in Allais's example, because the goal that can be attained by violating it is not as attractive. In Allais's example, the goal is $500,000. Here the goal is only a known (in the case of the pessimists) or unknown (in the case of the optimists) chance at a certain amount of money.

TABLE 6
The probabilities in Allais's example

	Probability of $2,500,000	Probability of $500,000	Probability of nothing
f	0.00	1.00	0.00
g	0.10	0.89	0.01
f'	0.00	0.11	0.89
g'	0.10	0.00	0.90

TABLE 7
Ellsberg's example

	s	t	u
f	$0	$100	$100
g	$100	$0	$100
f'	$0	$100	$0
g'	$100	$0	$0

4.3.2 Is the Postulate Absolutely Convincing?

Morgenstern (1979, page 180) described the independence postulate as "absolutely convincing"; reasonable people will violate it only if they do not understand it or do not realize how it applies to the problem they are considering. This claim is sometimes buttressed by the observation that both experimental subjects and students in decision theory classes can be convinced to change their preferences to agree with the postulate (MacCrimmon, 1968).

The claim that reasonable people will conform to the independence postulate when they fully understand it can never be conclusively refuted. Any failure to conform can always be attributed to unreasonableness or lack of understanding. Some reasonable people have been convinced that the postulate is not absolutely convincing, however, by the experimental work

This content downloaded from 150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

of Slovic and Tversky (1974). After querying college students about their preferences in the examples of Allais and Ellsberg, these authors explained the independence postulate to those who had violated it, explained the arguments for violating it to those who had obeyed it, and then gave both groups an opportunity to change their preferences. They also studied the effect of this information when it was presented before students were asked to express their preferences. They found that the arguments for violating the postulate were at least as persuasive as the arguments for obeying the postulate.

4.3.3 The Sure Thing Principle

Savage derived the independence postulate from a more intuitive but less precise principle that he called "the sure thing principle." Suppose $A_1, \ldots, A_n$ form a partition of the set S of states of a small world, and suppose f and g are acts. Suppose you are able to compare the consequences of f and g separately for each A_i, in abstraction from their consequences for the other A_i. You are able, that is to say, to say whether you prefer the pattern of consequences $\{f(s)\}_{s \in A_i}$ to the pattern of consequences $\{g(s)\}_{s \in A_i}$. The sure thing principle says that if you prefer $\{f(s)\}_{s \in A_i}$ to $\{g(s)\}_{s \in A_i}$ for each A_i, then you should prefer f to g.

This principle cannot itself serve as a postulate within Savage's system, because that system talks only about preferences between acts, not about preferences between partial acts such as $\{f(s)\}_{s \in A_i}$ and $\{g(s)\}_{s \in A_i}$. (An act is a mapping from S to C, not a mapping from just part of S to C.) But it seems more immediately understandable and appealing than the independence postulate.

The sure thing principle is appealing because it reflects a familiar strategy for resolving decision problems. When we are trying to decide what to do, we often devise a set $A_1, \ldots, A_n$ of mutually exclusive and jointly exhaustive situations and look for an act that seems to be advantageous or at least satisfactory in all these situations.

We cannot expect that this strategy will always be successful, however. It will not always produce a good argument, and even when it does, this argument may be outweighed by other arguments, as it is in Allais's example.

The strategy suggested by the sure thing principle may fail in several different ways. It may fail because we are unable to construct a convincing argument for any particular act when we consider a given A_i in isolation. It may fail because consideration of the different A_i may produce convincing arguments for different acts. Or, as in Allais's example, it may fail because there is a convincing argument that we will overlook when we consider the different A_i separately.

One reason it is sometimes difficult to construct a convincing argument for a particular act when we consider a given A_i in isolation is that the instructions "suppose you knew that the true state of the small world is in A_i" may not suffice to define a situation for us. (This is also part of the difficulty in turning the principle into a formal postulate.) In examples such as Allais's where objective chances are supplied for each state of the small world, there is an implicit message about how we should define this situation: we should renormalize these chances for the states s in A_i so they add to one. We see this in Savage's own discussion of Allais's example, where he writes of "a 10-to-1 chance to win \$2,500,000" (*Foundations*, page 103). But in problems where probabilities are not given *ex ante*, this solution is not available, and to assume that there are subjective probabilities available for renormalization begs one of the questions that Savage's postulates are supposed to resolve.

4.3.4 The Mixing Argument

The following argument in favor of the independence postulate has been used very effectively by Raiffa (1961, 1968). We may call it the mixing argument.

Suppose f, g, f', and g' satisfy the hypotheses of the independence postulate: $f'_A = f_A$, $g'_A = g_A$, $f_{A^c} = g_{A^c}$, and $f'_{A^c} = g'_{A^c}$. Suppose you violate the postulate by preferring f to g and g' to f'. Imagine I am about to toss a fair coin, and I offer you an opportunity to play the following compound game. If the coin comes up heads, then I will give you a choice between f and g. If the coin comes up tails, I will give you a choice between f' and g'. Since you prefer f to g and g' to f', you can tell me in advance what your choices will be. If the coin comes up heads, you will choose f; if it comes up tails you will choose g'. Let us call this your strategy: f if heads, g' if tails. The opposite strategy, which you apparently find less attractive, is g if heads, f' if tails.

But is there really anything to choose between these two strategies? If we let s denote the true state of nature, then your strategy gives you a 50–50 chance at $f(s)$ or $g'(s)$. The opposite strategy would give you a 50–50 chance at $g(s)$ or $f'(s)$. But these two 50–50 chances boil down to the same thing, no matter what s is. To see this, recall that (i) if $s \in A$, then $f(s) = f'(s)$ and $g(s) = g'(s)$, and (ii) if $s \in A^c$, then $f(s) = g(s)$ and $f'(s) = g'(s)$.

It is embarrassing enough that your preferences for f over g and g' over f' lead to a preference between two equivalent strategies, but things get worse. If you feel strongly about your preferences for f over g and g' over f', then presumably these preferences will not change when g and f' are both improved slightly. And the argument just given then shows that you prefer one strategy to another which is clearly better.

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

TABLE 8
The result of mixing your choices

	s	t	u
Your strategy	\$0 or \$500,000	\$500,000 or \$2,500,000	\$0 or \$500,000
The opposite	\$1 or \$500,001	\$500,001 or \$2,500,001	\$1 or \$500,001

Note: Each entry represents a 50–50 chance.

Just to make this last point as vivid as possible, let us rehearse it using Allais's example. Suppose your preferences for f over g and g' over f' are strong enough that they do not change when we increase all the entries for g and f' in Table 5 by \$1. Table 8 gives the results, in this case, of your strategy (f if tails, g' if heads) and the opposite strategy (g if heads, f' if tails). (All the entries in Table 8 should be interpreted as 50–50 chances; "\$0 or \$500,000," for example, means a 50% chance at \$0 and a 50% chance at \$500,000.)

This argument for the independence postulate can be persuasive, but a little thought will convince us that it is simply another way of deriving the postulate from the sure thing principle. The crucial step in the argument is the step where it is concluded from your preferences for f over g and g' over f' that you would prefer the strategy "f if heads, g' if tails" over the strategy "g if heads, f' if tails." This step can only be justified by appeal to the sure thing principle. Attention has shifted to a small world whose two states are heads and tails. Call heads B and tails B^c. The sure thing principle says that if you prefer the first strategy to the second when B is considered in isolation and also when B^c is considered in isolation, then you will prefer the first strategy to the second overall. But we need not obey this principle. We may refuse to do so on the grounds that our argument for choosing f over g—the fact that f guarantees us \$500,000—is not available when we must choose one of the strategies in Table 8.

4.3.5 The Imaginary Protocol

Another argument for the independence postulate, which has also been used effectively by Raiffa (1968, pages 82 and 83), asks us to imagine a protocol under which we find out about the true state of the small world in steps and do not have to choose between f and g or between f' and g' until after we have found out whether the state is in A.

Let us explain the argument, as Raiffa does, in terms of Allais's example, given in Table 5, with the states s, t, and u assigned probabilities .01, .10, and .89, respectively. Imagine that the determination of the true state of the small world is made by a two-stage random drawing. First you draw a ball from an urn containing 89 orange balls and 11 white balls. If you draw orange, then u is the true state of the small world. If you draw white, then you make a second drawing from an urn containing 10 red balls and one blue ball. If you draw the blue ball, then s is the true state; if you draw a red ball, then t is the true state.

Suppose you are asked to choose between f and g, but you are not required to do so until after the first drawing. If the first ball drawn is orange, then u is the true state, and you get \$500,000 in either case, so there is really no need to choose. But if the first ball drawn is white, then you are required to choose between f and g before making the second drawing.

This situation is depicted in Figure 1 (adapted from Raiffa, 1968, page 82). Notice that if the first drawing produces an orange ball, then you are awarded \$500,000 with no further ado; no choice or second drawing is required. Similarly, if the first drawing produces a white ball and you choose f, then you are awarded \$500,000 with no further ado; the second drawing is not required.

Figure 1 represents the choice between f and g. But with one simple change, it becomes a representation of the choice between f' and g'. We simply change the underlined \$500,000 to \$0.

The argument for the independence postulate now proceeds in three steps.

Step 1. Under the conditions just described, where you make your choice only if and when a white ball has been drawn, if you choose f over g, then you should also choose f' over g'. The situation where you must choose between f' and g' differs from the situation where you must choose between f and g only in what you *would have* received had you drawn an orange ball instead of a white ball. And surely you will want to base your choice on your present situation, not on might have beens.

Step 2. It should not make any difference if you are required to choose at the outset rather than only if and when you draw a white ball. You can mentally put yourself in the situation where you have just drawn a white ball, you know that in this situation you will prefer f to g and f' to g', and you know that only if you are later in this situation will the choice make any difference. So surely you should prefer f to g and f' to g' now.

Step 3. The choice between f and g or between f' and g' should only depend on the probability distributions of the consequences of these acts. So the conclusion, that if you prefer f to g then you should also prefer f' to g', must hold whenever the true state of the small world is s, t, or u, with probabilities 0.01, 0.10, and 0.89, respectively, even in the absence of the step by step protocol depicted in Figure 1.

A different premise is invoked at each step of this argument. In Step 1, present choices should not de-

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

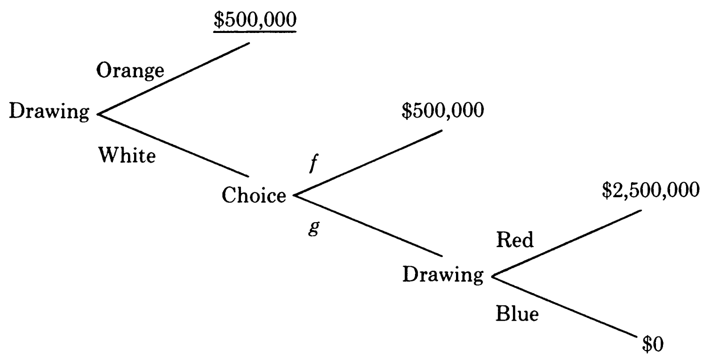

FIG. 1. *The imaginary protocol.*

pend on might have beens. In Step 2, if under the only scenario where a choice makes any difference there is a point at which you would choose in a certain way, that is also the way you should choose now. In Step 3, choices should depend only on the overall probability distributions of advantages and disadvantages, not on any protocol for the timing of your knowledge and choices.

I contend that none of these premises are compelling. They would be compelling if we could pretend that preferences are pre-existent and well-defined for every situation. But they are not compelling if we recognize that preferences are constructed.

The premise in Step 3 is especially objectionable, because it unreasonably limits the way in which preferences may depend on opportunities to adopt feasible goals. Consider the person who prefers g' to f' in the absence of the protocol depicted in Figure 1, but who would choose the $500,000 were he in the situation where a white ball has just been drawn. If there is no protocol, then he can argue that since he is likely to win nothing he might as well gamble with his slim chances. But if the protocol in Figure 1 is followed, and if he has just drawn a white ball, then he is in a position where $500,000 is a feasible goal. Why must he ignore this fact? (Even when we are concerned only with probability judgment and not with choice and preference, the presence or absence of a protocol is not irrelevant. See Shafer, 1985.)

The other two premises are also unpersuasive as general and apodictic principles, although they may be persuasive in particular cases. The first premise, which says that might have beens should not matter to you, overlooks the fact that your present preferences may be the result of goals that you adopted earlier, when what is now a might have been was a real possibility to you. Once we admit that goals and preferences are adopted or constructed, we cannot pretend that history is irrelevant. The second premise is open to the same objection, for it tries to rule out the adoption of any goal that might reverse the preference that you might guess your future self would have were you not to adopt the goal.

In the preceding discussion, I have focused on the version of Allais's example where probabilities are assigned *ex ante* to the possible states of the small world. More difficulties arise if one attempts to extend Raiffa's argument to everyday problems, where the state of one's small world is not determined by a chance device. In such problems it may be difficult to construct an imaginary protocol in which the first step leads to knowledge that the true state is or is not in A and no more; "now you know that the true state is in A" may not suffice to define a situation in which we can imagine ourselves.

4.3.6 Goals and Commodities

I have repeatedly used the idea of adopting goals to defend violations of the independence postulate. My point is that the process of formulating and adopting goals creates a dependence of value on belief, simply because goals are more attractive when they are feasible.

The dependence of the goal formation process on belief is the most fundamental reason for an adherent of the constructive view to reject the sure thing principle. The formation of goals does not usually take place at the level of individual states or restricted sets of states. Typically, we adopt goals that relate to the overall situation we are in. The adoption of goals ties states together, for the attractiveness of a goal depends on its meaningfulness and feasibility in all the states we consider possible, or at least in all the ones we consider probable.

It is interesting, in this connection, to recall the contrast between small worlds and commodities, drawn sharply by Samuelson (1952). Samuelson, who was at first reluctant to accept Savage's sure thing principle, finally did so because he became convinced that a person cannot make trade offs between small worlds the way he or she can make trade offs between commodities.

Suppose we want to rank in preference situations in which we have different amounts of three commodities—flour, apples, and butter. Let (x, y, z) denote the

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

situation where we have x pounds of flour, y apples, and z pounds of butter. Set

$$f = (4, 3, 1), \qquad f' = (4, 3, 0),$$
$$g = (2, 6, 1), \qquad g' = (2, 6, 0).$$

We may very well prefer f to g but g' to f'. If we had a pound of butter, we could make better bread, and so we would rather have more flour and fewer apples; this is a reason to prefer f to g. But if we do not have any butter, then flour is less interesting; we may prefer g' to f'. As this example illustrates, the amount of one commodity we have may influence the trade offs we make between other commodities. We cannot consider separately our preferences for the commodities in the disjoint sets A = {flour, apples} and A^c = {butter}, because what we can get in A^c influences our preferences within A. The goal of a loaf of bread ties A and A^c together.

Samuelson's conversion to the sure thing principle was based on the feeling that states of small worlds are not like commodities in this respect. Different states of small worlds are completely separate from one another. What we would have in one state of a small world cannot help us enjoy or use something we would have in another state. So we should be able to think separately about our preferences in disjoint sets of states. What we can get in A^c should not influence our preferences within A.

The constructive view forces us to recognize, however, that our value resides where it is constructed. If we construct goals within the products of our imagination that Savage called states of a small world, then the sure thing principle will hold. But if we construct goals in our real situation, then these goals may tie the states of the small world together as effectively as the goal of a loaf of bread ties butter and flour together.

5. THE PROBLEM OF SMALL WORLDS

A subjective expected utility analysis of a decision problem using one small world may fail to give the same result as an analysis using a more detailed small world. This is the problem of small worlds. As I pointed out in the introduction, this problem appears from the constructive viewpoint as just one more aspect of the lack of invariance of preference. The preferences we construct depend on the questions we ask ourselves, and hence the selection of questions is an essential part of the construction.

There is much to learn, however, from a closer look at the problem of small worlds. We can learn something about the nature of value from the very fact that we sometimes assign values to consequences at a limited level of description, without considering probabilities for further contingencies that might affect our enjoyment of these consequences.

In this section, I emphasize that the constructive view forces us to take seriously the fact that we work at limited levels of description. When we take the constructive view, we cannot pretend that every utility is really, at a more detailed level of description, an expected utility.

We can also profit from a closer look at the mathematical structure required to make one small world a refinement of another. In this section, I will describe this structure using a notation somewhat different from Savage's own. I will then develop a detailed example of two related small worlds. This is something Savage did not do, and by doing it we gain some insights he may have missed.

I conclude this section with a look at how Savage himself saw the problem of small worlds. For him, the problem was that refinement might change the probabilities that can be deduced from a person's preferences. The fact that Savage construed the problem of small worlds in this way demonstrates just how hopelessly nonconstructive his normative viewpoint was.

5.1 Are All Utilities Really Expected Utilities?

According to the constructive viewpoint, the method of subjective expected utility involves constructing preferences from separate judgments of value and belief. We distinguish between states of the world, about which we have evidence and for which we can construct probabilities, and consequences, to which we decide to attach values, represented numerically by utilities. The virtue of the method is that it breaks our deliberation into simpler and more manageable parts. We can deal separately with evidence for which state of the world is true and arguments about what we should value.

The idea of refinement threatens this picture. A thing to which we might want to assign a definite value at one level of description seems, at a finer level, to have a value that depends on how various questions of fact turn out. It seems that every utility, on closer examination, is an expected utility.

In the preceding section I pointed out that a subjective expected utility argument requires a small world (S, C) such that our preferences over C do not depend on which description in S is true. I argued that sometimes we will be unable to make a subjective expected utility argument because we are unable to devise such a small world. Here I am raising a related but different point. I am considering the case where we do make a subjective expected utility argument—the case where we do succeed in devising a small world (S, C) such that we are willing to settle on preferences over C that are independent of which description in S is true—and I am asking whether these preferences might still depend on questions of fact more detailed than those answered by the descriptions in S.

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

Savage discussed this point as follows on pages 83 and 84 of *Foundations*:

> . . . Jones is faced with the decision whether to buy a certain sedan for a thousand dollars, a certain convertible also for a thousand dollars, or to buy neither and continue carless. The simplest analysis, and the one generally assumed, is that Jones is deciding between three definite and sure enjoyments, that of the sedan, the convertible, or the thousand dollars. Chance and uncertainty are considered to have nothing to do with the situation. This simple analysis may well be appropriate in some contexts; however, it is not difficult to recognize that Jones must in fact take account of many uncertain future possibilities in actually making his choice. The relative fragility of the convertible will be compensated only if Jones's hope to arrange a long vacation in a warm and scenic part of the country actually materializes; Jones would not buy a car at all if he thought it likely he would immediately be faced by a financial emergency arising out of the sickness of himself or of some member of his family; he would be glad to put the money into a car, or almost any durable goods, if he feared extensive inflation. This brings out the fact that what are often thought of as consequences (that is, sure experiences of the deciding person) in isolated decision problems typically are in reality highly uncertain. Indeed, in the final analysis, a consequence is an idealization that can perhaps never be well approximated . . .

When we first consider an example like this one, we are tempted to think that a sufficiently detailed description of Jones's possible future situations would make it possible for him to decouple his utilities from his probabilities. But as I argued in Section 4.1, this sufficiently detailed description is a chimera. No matter how much detail we include in a description of a situation, there always remain uncertainties that can affect the degree to which we will enjoy or value that situation. This is the point of Savage's last sentence.

When we take a constructive view, we can no longer pursue the chimera of the sufficiently detailed description. Instead, we are forced to take seriously the idea that when a person decides to attach a value or utility to a consequence described at a certain limited level of detail, he or she does this and nothing more.

"I have decided to buy a convertible," Jones tells me, "because my wife and I are taking a vacation to New Mexico this summer, and we really want to enjoy the sun." "You should think this through more carefully, Jones," I respond. "Don't you remember that sunburn you got at Daytona Beach last spring? You never really enjoy these vacations anyway. And if your wife does like the sun that much, she may not come back to Chicago with you." "You are always dreaming up things to worry about," replies Jones. "I detest this winter weather, and I have set my heart on a tour of the desert in the sun. The trip may be a disaster, but staying home might be a disaster, too. Who knows?"

Jones has decided on a trip to sunny New Mexico in a convertible. He does not want to analyze all the different ways taking the trip might turn out, partly because he does not feel he can construct convincing probabilities for them, but also because these more detailed scenarios are not really the objects of his desire. The trip lies within the bounds of prudent behavior, and he and his wife have decided they want to go.

A constructive interpretation of subjective expected utility must hold that a utility is not an expected utility in disguise. A utility is a value deliberately attached to a consequence created at a given level of description. The consequence is a product of our imagination. The utility is a product of our will. We may later analyze the consequence at a finer level of description, and we may then assign it an expected utility rather than just a utility. But any such further analysis is a further act of imagination and will, not something already determined or achieved.

5.2 Refining Small Worlds

In order to study the problem of small worlds as it appears from Savage's normative point of view, we need to understand the technical aspects of refining a small world. Suppose (S, C) and (T, D) are two small worlds. How do we give mathematical form to the idea that (T, D) is a refinement of (S, C)?

Savage answered this question on pages 84–86 of *Foundations*. Unfortunately, he did so in the context of a "tongue in check" (page 83) assumption that (T, D) is actually a "grand world," i.e., an ultimately detailed refinement. This assumption does not affect the technical details of the mathematical structure relating (S, C) and (T, D), but it did, I think, obscure Savage's view. Since he was struggling with the idea of (T, D) being a grand world, he missed the insight he might have gained from a concrete example where (S, C) and (T, D) are both small worlds. (The only example he gave was purely mathematical.) Moreover, he was content to "hobble along" (page 85) with an inadequate notation.

When we examine Savage's account, on pages 84 and 85 of *Foundations*, we see that (S, C) and (T, D) are related in two ways. First, the descriptions in T are more detailed versions of the descriptions in S. Second, the consequences in C correspond to acts in (T, D), mappings from T to D. It is easy to establish a mathematical notation for both these aspects of the relation. For each element t of T, let t^* denote the

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

unique element of S that agrees with t but is less detailed. And for each element c of C, let c^* denote the corresponding act in (T, D).

Why does a consequence in the less refined small world (S, C) correspond to an act in the more refined small world (T, D)? We might think that just as the descriptions of states of the world in T are merely more detailed versions of the descriptions in S, so the descriptions of the states of the person in D should merely be more detailed versions of the descriptions in C. But Savage felt that pushing to a more refined level of description may mean more than describing the same consequences in more detail. It may mean instead shifting attention to entirely different and more fundamental consequences. We may, for example, shift our attention from monetary income to personal satisfaction. The same level of satisfaction can be achieved with different levels of income, depending on the state of the world. So if the elements of C are levels of income, and the elements of D are levels of satisfaction, then we do not want to say that each element of D is a more detailed version of some element of C. Instead, we want to say that each element of C determines an element of D when combined with a state of the world in T. This can be expressed mathematically by saying that each element of C corresponds to a mapping from T to D.

Once we have linked (S, C) and (T, D) by specifying t^* for each t in T and c^* for each c in C, we also have a way of relating acts in (S, C) to acts in (T, D). Suppose, indeed, that f is an act in (S, C). Then there is a unique act in (T, D), say f^*, that corresponds to f. The act f^* maps a given element t of T to $(f(t^*))^*(t)$, which is an element of D.

It is interesting and important to note that this mathematical apparatus linking (S, C) and (T, D) goes beyond what we construct when we formulate two small worlds in two different attempts to study the same set of concrete acts. In order to see this clearly, we need a notation that distinguishes between a concrete act and its representation as a mapping within a particular small world. Given a concrete act a, let f_a^S denote the coresponding abstract act in (S, C), and let f_a^T denote the corresponding abstract act in (T, D); f_a^S is a mapping from S to C, and f_a^T is a mapping from T to D.

Suppose we formulate (S, C) and (T, D) in separate attempts to construct a setting for deliberation about a set F_0 of concrete acts. We formulate (S, C) first, find it too crude, and then formulate (T, D) in order to deepen our analysis. This exercise results in four sets of written descriptions, S, T, C, and D, and mappings f_a^S and f_a^T for each concrete act a in F_0. Since S and T consist of written descriptions, the relation between them will be clear; for each t in T, we will be able to pick out t^*, the unique element of S that agrees with t but is less detailed. Moreover, since we have identified the concrete acts in F_0 with acts in (S, C) and (T, D), we have partially determined mappings c^* corresponding to the elements c of C. We may not have fully determined these mappings, however. We must have $(f_a^S)^* = f_a^T$ for all a in F_0. Equivalently, we must have

$$(f_a^S(t^*))^*(t) = f_a^T(t)$$

for all a in F_0 and all t in T. This determines $c^*(t)$ whenever there is a concrete act a that results in the consequence c if t^* is the true state of the small world (S, C). But there are usually pairs (c, t) for which there is no such concrete act a, and $c^*(t)$ will not be determined for these pairs.

Consider an example. Begin with the small world (S, C) given in Table 1 in Section 2.1. This is the small world that Savage formulated for the omelet maker who must decide whether to crack a sixth egg into a bowl already containing five eggs. Suppose the omelet maker decides to refine (S, C) because he realizes that his guests can distinguish between a Nero Wolfe omelet, i.e., one made with eggs less than 36 hours old, and an ordinary omelet, i.e., one made with eggs that are not so fresh. He refines the states of the world to take the freshness of the eggs into account, and he refines the consequences to take the quality of the omelet into account. Suppose, for simplicity, that the person knows that the five eggs in the bowl are all of similar freshness, and that the sixth egg, if it is good, will not affect whether the omelet meets Nero Wolfe standards. In this case we can use a set T consisting of four states of the world,

$$T = \left\{ \begin{array}{l} \text{the sixth egg is good,} \\ \quad \text{and the other five are fresh} \\ \text{the sixth egg is good,} \\ \quad \text{and the other five are stale} \\ \text{the sixth egg is rotten,} \\ \quad \text{and the other five are fresh} \\ \text{the sixth egg is rotten,} \\ \quad \text{and the other five are stale} \end{array} \right\},$$

and a set D consisting of the eleven consequences listed in Table 9. We are still considering the same concrete acts:

$$F_0 = \left\{ \begin{array}{l} \text{break the egg into the bowl} \\ \text{break the egg into a saucer} \\ \text{throw the egg away} \end{array} \right\}.$$

Table 9 shows how the three acts map T to D.

When we compare Tables 1 and 9, we see that these tables determine some of the values $c^*(t)$ but not others. Consider, for example, the first of the three consequences in Table 1, "six-egg omelet." For brevity,

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

TABLE 9
A refinement of Savage's small world

Act	State: Good, Fresh	State: Good, Stale	State: Rotten, Fresh	State: Rotten, Stale
Break into bowl	Six-egg Nero Wolfe omelet	Six-egg ordinary omelet	No omelet and five good eggs destroyed	No omelet and five good eggs destroyed
Break into saucer	Six-egg Nero Wolfe omelet and a saucer to wash	Six-egg ordinary omelet and a saucer to wash	Five-egg Nero Wolfe omelet and a saucer to wash	Five-egg ordinary omelet and a saucer to wash
Throw away	Five-egg Nero Wolfe omelet and one good egg destroyed	Five-egg ordinary omelet and one good egg destroyed	Five-egg Nero Wolfe omelet	Five-egg ordinary omelet

let this consequence be denoted by c_1. It is clear that

$$c_1^*(\text{good, fresh}) = \text{six-egg Nero Wolfe omelet},$$

and

$$c_1^*(\text{good, stale}) = \text{six-egg ordinary omelet}.$$

The six-egg omelet is of Nero Wolfe or ordinary quality depending on whether the five eggs are fresh or stale. But what are c_1^*(rotten, fresh) and c_1^*(rotten, stale)? If, by magic, we get a six-egg omelet even though the sixth egg is rotten, then what is the quality of this six-egg omelet? This question is not answered by Tables 1 and 9. We may be inclined to say that the six-egg omelet will still be a Nero Wolfe omelet if the five eggs are fresh and an ordinary omelet if the five eggs are stale, but this is not a statement of fact. It is merely a natural way to exercise our imagination. We could invent other examples where it is more difficult to settle on a natural way of exercising our imagination.

Savage seems to have overlooked this remarkable extent to which the structure relating small worlds is a product of our imagination, perhaps because he did not study conrete examples. Perhaps this contributed to his reluctance to see any force in the objections to his use of imaginary acts (Section 3.4). It seems reasonable to put an imaginary act that always yields a six-egg omelet into our preference ranking if we permit ourselves to think of a six-egg omelet as a simple object of desire. It seems less reasonable if the very meaning of a six-egg omelet depends on deliberate and not yet performed acts of imagination.

5.3 Savage's Problem of Small Worlds

Consider a small world (S, C) and a refinement (T, D). Suppose a person has preferences over acts in (T, D) that satisfy Savage's postulates and hence determine a probability measure P_T on T and a utility function U_D on D. From these preferences, probabilities, and utilities, how do we find the person's probability measure P_S and utility function U_C for (S, C)? There are two possible methods.

Method 1. Since S amounts to a disjoint partition of T, we can take P_S to be P_T's marginal on that disjoint partition. And we can say that the person's utility for a consequence c in C is his expected utility for the act c^* in (T, D); $U_C(c) = E_T(U_D(c^*))$.

Method 2. Since every act f in (S, C) can be identified with an act f^* in (T, D), the person's preferences over acts in (T, D) determine preferences over acts in (S, C). If these letter preferences satisfy Savage's postulates, then they directly determine a probability measure P_S and a utility function U_C.

For Savage, the problem of small worlds was that these two methods may fail to produce the same answer. Savage showed that if the preferences over (S, C) do satisfy his postulates, so that Method 2 is applicable, then the two methods will give the same utility function on C. But they may give different probability measures on S (*Foundations*, pages 88–90).

I will not reproduce Savage's mathematical reasoning here. But I will illustrate the problem using the example of the omelet. Suppose a person's preferences over the small world (T, D) of Table 9 satisfy Savage's postulates and yield the probabilities and utilities shown in Table 10. (In order for Savage's sixth postulate to be satisfied and the probabilities and utilities to be fully determined, we would need to refine T further so that each state specifies the outcome, say, of a sequence of coin tosses. But we need not make such a refinement explicit here.) According to the probabilities in Table 10, the sixth egg is as likely to be rotten as good, but its being good makes it more likely that the other five are fresh. The utilities indicate that the person is indifferent as to whether or not he washes a saucer or destroys a good egg, but that he prefers a six-egg omelet to a five-egg one and a Nero Wolfe omelet to an ordinary one. (Since the person is indifferent about washing the saucer or destroying a good egg, Table 10 omits these details in assigning utilities to the consequences in D. It is to be

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

understood, for example, that the person assigns utility 8 to both "five-egg ordinary omelet" and "five-egg ordinary omelet and a saucer to wash.")

The probabilities and utilities that Table 10 gives for (T, D) result in preferences among the acts in the smaller world (S, C) that do satisfy Savage's postulates, and so we can use both Method 1 and Method 2 to obtain probabilities and utilities for (T, D). The results are shown in Table 11. Only one set of utilities is given in this table; as we have mentioned, Savage showed that when Method 2 is applicable it necessarily gives the same utilities as Method 1. (In Table 11, as in Table 10, the consequences are described only in the relevant degree of detail.) But the two methods give different probabilities for the sixth egg's being good.

The reader can easily check the numbers given in Table 11 for Method 1. To obtain the probability of the sixth egg being good, add the probabilities 3/8 and 1/8 from Table 10. To obtain the expected utility of a five-egg omelet, calculate

$$P_T(\text{fresh})U_D(\text{five-egg Nero Wolfe omelet}) + P_T(\text{stale})U_D(\text{five-egg ordinary omelet}) = (5/8)(16) + (3/8)(8) = 13.$$

And so on.

According to Method 2, the probability of the sixth egg being good is more than 1/2. Why? Because an omelet is valued more highly when the eggs are fresh than when they are stale. The distinction between fresh and stale cannot be expressed in (S, C), but since the five eggs are more likely to be fresh when the sixth is good, the preference for fresh over stale shows up as a preference for an act that gives an omelet when the sixth is good over an act that gives an omelet when the sixth is rotten. This gives the impression that the person puts a higher probability on its being good.

How do we get the exact value 7/13 for $P_S(\text{good})$? One way is to apply formula (7) on page 88 of *Foundations*. A quicker way is to equate $E_S(f)$ and $E_T(f^*)$ for the act f in (S, C), where

$$f(\text{good}) = \text{five-egg omelet}, \quad f(\text{rotten}) = \text{no omelet}.$$

We have

$$E_S(U_C(f)) = P_S(\text{good}) \times 13 + P_S(\text{rotten}) \times 0 = P_S(\text{good}) \times 13.$$

And since

$$f^*(\text{good, fresh}) = \text{five-egg Nero Wolfe omelet},$$
$$f^*(\text{good, stale}) = \text{five-egg ordinary omelet},$$
$$f^*(\text{rotten, fresh}) = \text{no omelet},$$
$$f^*(\text{rotten, stale}) = \text{no omelet},$$

we have

$$E_T(U_D(f^*)) = P_T(\text{good, fresh})U_D(\text{five-egg Nero Wolfe omelet}) + P_T(\text{good, stale})U_D(\text{five-egg ordinary omelet}) + P_T(\text{rotten, fresh})U_D(\text{no omelet}) + P_T(\text{rotten, stale})U_D(\text{no omelet}) = (3/8)(16) + (1/8)(8) + (1/4)(0) + (1/4)(0) = 7.$$

Equating the two expected values, we obtain $P_S(\text{good}) = 7/13$.

The possible divergence between Methods 1 and 2 disturbed Savage. He was not disturbed by the possibility that preferences over acts in a small world may fail to satisfy his postulates, for this can be taken as a signal that the small world needs to be refined. But he was disturbed by the possibility that probabilities calculated in a small world that did satisfy his postulates might change with refinement. If the probabilities are different for two different levels of refinement, then which level is right? How can we tell?

Savage posited the existence of a grand world in order to answer the first of these two questions. Probabilities calculated from a given small world are right if they are the same as the ones calculated from the grand world. Yet even this outrageous fiction left him without an answer to the second question. How can we tell if the probabilities from a given small world are the same as the ones we would get if, counter to fact, we were able to work with a grand world?

TABLE 10
Probabilities and utilities for (T, D)

States	Probabilities	Consequences	Utilities
Good, fresh	3/8	No omelet	0
Good, stale	1/8	Five-egg ordinary omelet	8
Rotten, fresh	1/4	Five-egg Nero Wolfe omelet	16
Rotten, stale	1/4	Six-egg ordinary omelet	16
		Six-egg Nero Wolfe omelet	32

TABLE 11
Probabilities and utilities for (S, C)

States	Probabilities		Consequences	Utilities
	Method 1	Method 2		
Good	1/2	7/13	No omelet	0
Rotten	1/2	6/13	Five-egg omelet	13
			Six-egg omelet	26

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms

Savage called a small world that satisfied his postulates a *pseudomicrocosm.* He called a pseudomicrocosm which would give the same probabilities as the grand world a *real microcosm.* He wrote, "I feel, if I may be allowed to say so, that the possibility of being taken in by a pseudomicrocosm that is not a real microcosm is remote, but the difficulty I find in defining an operationally applicable criterion is, to say the least, ground for caution" (*Foundations*, page 90).

The possibility of being taken in by a pseudomicrocosm that is not a real microcosm is indeed remote. It is remote because one could not possibly have detailed preferences among acts satisfying Savage's postulates unless one deliberately constructured these postulates from probabilities and utilities. Thus Savage's version of the problem of small worlds serves as a demonstration of how far his normative approach was from a sensible, constructive approach to decision.

ACKNOWLEDGMENTS

Research for this article has been partially supported by Grants MCS-800213 and IST-8405210 from the National Science Foundation. The author has benefited from conversation and correspondence with Morris DeGroot, Peter Fishburn, Dennis Lindley, Pamela Townsend, and Amos Tversky.

REFERENCES

ALLAIS, M. (1953). Fondements d'une theorie positive des choix comportant un risque et critique des postulats et axiomes de l'école Americaine. *Colloques Internationaux du Centre National de la Recherche Scientifique, Econometrie* **40** 257–332. (English translation with the title, "The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American school," in Allais and Hagen, 27–145, 1979.)

ALLAIS, M. (1979). The so-called Allais paradox and rational decisions under uncertainty. In Allais and Hagen, 437–681, 1979.

ALLAIS, M. and HAGEN, O., eds. (1979). *Expected Utility Hypotheses and the Allais Paradox.* Reidel, Dordrecht.

ANSCOMBE, F. J. (1956). Review of *The Foundations of Statistics* by Leonard J. Savage. *J. Amer. Statist. Assoc.* **51** 657–659.

AUMANN, R. J. (1962). Utility theory without the completeness axiom. *Econometrica* **30** 455–462.

AUMANN, R. J. (1964). Utility theory without the completeness axiom: a correction. *Econometrica* **32** 210–212.

ELLSBERG, D. (1961). Risk, ambiguity, and the Savage axioms. *Quart. J. Econom.* **75** 643–669.

FISCHHOFF, B., SLOVIC, P. and LICHTENSTEIN, S. (1980). Knowing what you want: Measuring labile values. In Wallsten, 117–141, 1980.

FISHBURN, P. C. (1970). *Utility Theory for Decision Making.* Wiley, New York. Reprinted by Krieger in 1979.

FISHBURN, P. C. (1981). Subjective expected utility: a review of normative theories. *Theory Decision* **13** 139–199.

FRIEDMAN, M. and SAVAGE, J. (1952). The expected utility hypothesis and the measurability of utility. *J. Polit. Econ.* **60** 464–474.

KAHNEMAN, D., SLOVIC, P. and TVERSKY, A., eds. (1982). *Judgment under Uncertainty: Heuristics and Biases.* Cambridge Univ. Press, Cambridge.

LINDLEY, D. V. (1974). Discussion of papers by Professor Tversky and Professor Suppes. *J. Roy. Statist. Soc. Ser. B* **36** 181–182.

LUCE, R. D. and KRANTZ, D. H. (1971). Conditional expected utility. *Econometrica* **39** 253–271.

MACCRIMMON, K. R. (1968). Descriptive and normative implications of the decision-theory postulates. In *Risk and Uncertainty* (K. Borch and J. Mossin, eds.) 3–32. Macmillan, New York.

MARSCHAK, J. (1950). Rational behavior, uncertain prospects, and measurable utility. *Econometrica* **18** 111–141.

MAY, K. O. (1954). Intransitivity, utility, and the aggregation of preference patterns. *Econometrica* **22** 1–13.

MORGENSTERN, O. (1979). Some reflections on utility. In Allais and Hagen, 175–183, 1979.

PRATT, J. W. (1974). Some comments on some axioms for decision making under uncertainty. In *Essays on Economic Behavior Under Uncertainty* (M. Balch, D. McFadden and S. Wu, eds.) 82–92. North-Holland, Amsterdam.

RAIFFA, H. (1961). Risk, ambiguity, and the Savage axioms: comment. *Quart. J. Econom.* **75** 690–694.

RAIFFA, H. (1968). *Decision Analysis: Introductory Lectures on Choices under Uncertainty.* Addison-Wesley, Reading, Mass.

RICHTER, M. K. (1975). Rational choice and polynomial measurement models. *J. Math. Psych.* **12** 99–113.

SAMUELSON, P. A. (1952). Probability, utility, and the independence axiom. *Econometrica* **20** 670–678.

SAVAGE, L. J. (1954). *The Foundations of Statistics.* Wiley, New York. Second edition published by Dover in 1972.

SAVAGE, L. J. (1967). Difficulties in the theory of personal probability. *Philos. Sci.* **34** 305–310.

SAVAGE, L. J. (1971). Elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.* **66** 783–801.

SCHOEMAKER, P. J. H. (1982). The expected utility model: its variants, purposes, evidence and limitations. *J. Econom. Lit.* **20** 529–563.

SHAFER, G. (1981). Constructive probability. *Synthese* **48** 1–60.

SHAFER, G. (1985). Conditional probability. *Internat. Statist. Rev.* **53** 261–277.

SHAFER, G. and TVERSKY, A. (1985). Languages and designs for probability judgment. *Cog. Sci.* **9** 309–339.

SLOVIC, P. and TVERSKY, A. (1974). Who accepts Savage's axiom? *Behavioral Sci.* **19** 368–373.

TVERSKY, A. (1969). Intransitivity of preferences. *Psycholog. Rev.* **76** 31–48.

TVERSKY, A. (1972). Choice by elimination. *J. Math. Psych.* **9** 341–367.

TVERSKY, A. and KAHNEMAN, D. (1981). The framing of decisions and the psychology of choice. *Science* **211** 453–458.

TVERSKY, A. and KAHNEMAN, D. (1986). Rational choice and the framing of decisions. *J. Business* **59** S251–S278.

VON NEUMANN, J. and MORGENSTERN, O. (1944, 1947, 1953). *Theory of Games and Economic Behavior.* Princeton Univ. Press, Princeton, N.J.

WALLSTEN, T. S. (1980). *Cognitive Processes in Choice and Decision Behavior.* Erlbaum, Hillsdale, N. J.

WOLFOWITZ, J. (1962). Bayesian inference and axioms of consistent decision. *Econometrica* **30** 470–479.

This content downloaded from
150.214.205.43 on Mon, 25 Jan 2021 14:04:11 UTC
All use subject to https://about.jstor.org/terms