(2)

Giving Up Certainties DRAFT

1. Introduction.

People have worried for many years — centuries — about how you perform large changes in your body of beliefs. How does the new evidence lead you to replace a geocentric system of planetary motion by a heliocentric system? How do we decide to abandon the principle of the conservation of mass?

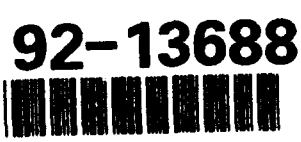
The general approach that we will try to defend here is that an assumption, presupposition, framework principle, will be rejected or altered when a large enough number of improbabilities must be accepted on be basis of our experience. If I think that all swans are white, and a student claims to have a counterexample, I will assume that he has made some observational error. I will reject his result, and continue to accept the generalization. When a lot of people claim to have seen counterexamples, I will come around: to continue to accept the generalization would require me to accept too many improbabilities. This is a discontinuous process as we will construe it: it is not a matter of a general statement becoming less probable, while certain reports become more probable. We cannot accept the generalization and even one of the observation reports: that would be a simple inconsistency.

One suggestion, due to Karl Popper, is that we invent Bold Conjectures, and Put Them to the Test. (Popper, the logic of scientific discovery) Bold conjecture: the Earth is the Center of the Solar System. Test... what? Bold conjecture: Mass is conserved. Test: weigh a mass of plutonium and its by products before and after. Obviously things are a little more complicated than the slogans suggest.

Alternatively, gather evidence, and accept the hypothesis that is most probable, relative to that evidence. So far, so good (maybe). But then what? How do you change from that hypothesis to one inconsistent with it when the evidence so indicates? For as soon as a hypothesis is accepted, it has probability 1; and as soon as a hypothesis has probability 1, its contraries have probability 0; and as soon as a contrary hypothesis has probability zero, its probability can never leave zero -- at least not by Bayes' theorem.

A natural response to this observation is to say (as Carnap did) "acceptance" is just an approximation to the real truth, which is that no hypothesis ever achieves literal acceptance, which would entail its having a probability of 1. What we really have (as opposed to the approximate way we talk) as a probability blanket over a field of empirical propositions, none of which is ever assigned a probability of 0 or 1 unless it is a mathematical or logical truth, or the denial of one.

This latter approach presents us with serious problems. We will consider the problem of assigning prior probabilities to the sentences of a reasonably rich language later, but already we are faced with a difficult computational problem. Gil Harman (Change in View) has pointed out that in a language win n basic sentences there are 2^n assignments to make. But of course we can get by with wholesale



assignments; if we decide that each conjunction of basic sentences or their negations is to have the same measure assigned to it, there is in fact only one assignment to make: one simple algorithm that provides the measure for any sentence.

In general, however, a useful and realistic language will have an infinite number of sentences, and this procedure breaks down. It is still possible to assign measures systematically, without assigning zero to any sentence representing a possibility. The number of sentences in any ordinary language is denumerable, and we can find a denumerable number of finite numbers that add up to 1. But the rationale of the system is hard to find.

It is, at any rate, worth exploring alternatives to either of these approaches to to rational acceptance. One of the first to offer a systematic procedure for this was Isaac Levi. In *Gambling with Truth* and *The Enterprise of Knowledge,* Levi proposes a rule for *adding* to your body of knowledge. Given such a rule, one can obtain a rule for replacing one conjecture, law, theory, hypothesis by another by proposing that when faced with a choice, one simply deletes both candidates from one's body of knowledge, and then adds the one indicated by the application of the rule for addition.

The rule is just this: [p.89] Let U be a set of most specific possible hypotheses — i.e., a set of which exactly one member is true. Let M be an "information determining probability" [Enterprise, p. 48],: M(g) represents the informational value of rejecting g, and let P be an expectation forming probability (a degree of belief, a credibility). let q in [0,1] be an index of caution. The rule (Rule A, of Gambling) is to reject all and only those elements g of U such that $P(g) < qM(g)$, and to accept, with deductive closure, the disjunction of the remainder.

Given a rule for acceptance, we can construe contraction as suspending belief in a proposition and then failing to add it back under subsequent expansion; and we can construe replacement as suspending belief in one proposition, and arriving at another on subsequent expansion.

We can accomplish a change of framework of "accepted facts" this way, and we can be sure of maintaining consistency in the process. There are some knotty problems, however. When and how do we decide to suspend belief in a framework proposition? There are clear cases: when observations render our corpus inconsistent, for example. "For the sake of argument," in a friendly social context. But the question has not been very thoroughly explored. How should q, the index of caution, be chosen? Where does the information measure M come from? How do we arrive at the credal probability p? More fundamental: How is the "abductive" step — the step in which the ultimate partition U is formed — to be controlled and rationalized?

One can always raise questions, of course. But these questions are disturbing because the rule presupposes a framework (a language, an information measure, a credibility measure, a set of most specific answers), and thus to be not even potentially capable of providing guidance in the choice of a framework. But let us look further.

An approach similar to Levi's has been developed in various ways by Makinson, Alchurron, and Gardenfors. While Levi approaches the question from a constructive, analytic angle, and seeks to provide formal analysis of what goes on in changes in a corpus of knowledge, Gardenfors and the others approach the question from a logical point of view: they seek to explore axioms that may be taken to characterize the change of a body of knowledge, construed as a set of propositions. Thus, for example, it is clear that if we add the proposition A to our body of knowledge K, then A should belong to that expanded body of knowledge. As is the case with Levi, it is assumed by these writers that a body of knowledge K should be construed as a deductively closed set of propositions.

An excellent examination of these logics of theory change is provided by Gardenfors' book, *Knowledge inf Flux*. It is from that source that I take the following axioms. A *belief set* here is construed as a deductively closed set of propositions.

If we denote by K_A^+ the expansion of a body of knowledge K by the addition of the consistent proposition A, then we may express the the properties of the expansion of a belief set by the following relatively uncontroversial axioms.

(K+ 1) K^+ is a belief set.

(K+ 2) $K_A^+ \supseteq A$

(K+ 3) If $\sim A \notin K_A^+$, then $K_A^+ \supseteq K$

(K+ 4) If $A \in K$, then $K_A^+ = K$

(K+ 5) If $H \supseteq K$, $H_A^+ \supseteq K_A^+$

(K+ 6) For all belief sets K and all sentences A, K_A^+ is the smallest belief set that satisfies (K+ 1) – (K+ 5).

What is not so uncontroversial is the question of the principles according to which a body of knowledge should be contracted. This is not a terribly serious question for Levi: any proposition in our body of knowledge can be doubted with relative impunity. It can be doubted with relative impunity, since, if it belongs in our corpus of knowledge, it will be reinstated on reflection. One can thus suspend belief in a proposition A on quite casual grounds.

A serious reason to suspend belief in something arises from the circumstance that our corpus of knowledge is inconsistent. For example, if there are observational routines that warrant our acceptance of the statement that *a* is a crow and *a* is not black, then when we practise those routines, we should accept the corresponding statement. (Or proposition.) But if we already accept the generalization that all crows are black, this renders our corpus inconsistent.

With an inconsistent corpus, we are clearly *obligated* to suspend belief in something. Levi says that we should shrink our corpus of knowledge in such a way as to retain the most "information." But it is clear that no simple-minded construal of "information" will lead to the right results. In some sense it is clear that the information content of "all crows are black" is greater than that of "*a* is a crow and *a* is not black," but of course on any standard construal of hypothesis testing it is the former that will be suspended and the latter that will be retained.

While Levi offers us no logic of contraction, that is the main concern of Gardenfors et al. Gardenfors offers a number of axioms characterizing the contraction operation, denoted by K^-_A, Most of these axioms are relatively uncontroversial, as in the case of expansion. We have:

(K⁻ 1) For any sentence A and any belief set K, K^-_A is a belief set.

(K⁻ 2) $K \supseteq K^-_A$.

(K⁻ 3) If $A \notin K$, then $K^-_A = K$.

(K⁻ 4) If not $\vdash A$, then $A \notin K$..

(K⁻ 5) If $A \in K$ then $K^-_A \supseteq K$.

(K⁻ 6) If $\vdash A \leftrightarrow B$, then $K^-_A = K^-_B$.

(K⁻7) $K^-_{A\&B} \supseteq K^-_A \cap K^-_B$

(K⁻8) If $A \notin K^-_{A\&B}$, then $K^-_A \supseteq K^-_{A\&B}$

These axioms may be more controversial than those for the expansion of a body of knowledge, but there is still nothing obviously wrong with them. It is possible to provide intuitively plausible axioms for theory replacement, and to show that in general replacement can be construed as a contraction followed by an expansion.

What becomes controversial is the procedure for conducting contractions. The contraction K_A^- is not uniquely determined by these axioms, in contrast to K_A^+ (under the assumption of deductive closure). We must thus consider *how* to perform the contraction. One possibility is the following. Consider a subset of K that is deductively closed, that does not contain A, and that is such that if any other sentence of K is added to it, A will be a consequence of it. The set of all such sets of sentences is denoted by $K \perp A$. Clearly the result of contraction should be a member of this set (if it isn't empty; if A is a theorem, then we can take the contraction of K by A to be K itself. All we need to do is to devise a "selection function" S that will pick one set out of $K \perp A$. But, as Gardenfors shows, this yields contractions that are "too big." If $A \in K$ then this procedure will yield a K_A^- that for any proposition B contains either $A \vee B$ or $A \vee \sim B$.

The next idea one might have is to consider the intersection of all the sets of sentences in $K \perp A$. (This is called the "full meet contraction.") This is too small: K_A^- will consist only of the logical consequences of $\sim A$.

Finally, we may consider a selection function S that picks *some* of the members of $K \perp A$, intuitively the most *epistemically entrenched* members, and take K_A^- to be the intersection of these.

But what does epistemic entrenchment come to? That seems to be where the real controversy lies. Levi seeks to preserve information (in some sense); he can be construed as construing epistemic entrenchment in terms of information. But the epistemic entrenchment ranking of sets of propositions can plausibly be taken to reflect a system of beliefs, and thus be sensitive to such things as "scientific revolutions." Gardinfors says that "...the fundamental criterion for determining the epistemic entrenchment of a sentence is how useful it is in inquiry and deliberation." [p.87] (Note that the selection function S is originally defined over sets of sentences, rather than sentences. This reflects a difference that could be exploited.

One idea for representing such factors is provided by Wolfgang Spohn ("Ordinal Conditional Functions: A Dynamic Theory of Epistemic States," in *Causation in Decision, Belief Change, and Statistics*, W. Harper and B Skyrms (eds) Reidel, Dordrecht, 1987.pp. 105-134.). Spohn defines an "ordinal conditional function" that maps possible worlds into ordinals. The value of the function represents a degree of implausibility, or a degree of unwillingness to accept, or a degree of potential surprise (Levi, Shackle).

This function is can be extended to propositions in general by taking the value of the function for a proposition, to be the minimum value of the function

over the set of worlds in which that proposition is true. Thus, since it is assumed that there is some world with value 0, either $k(A) = 0$ or $k(\sim A) = 0$, and $k(A \cup B) = \min\{k(A),k(B)\}$, where k is Spohn's ordinal conditional function.

Spohn's approach is more general than Gardenfors' since it takes as epistemic input a pair (A,a) consisting of a proposition A and an ordinal a. This yields a new ordinal function on possible worlds, and thus a new ordinal function. In the extreme cases, however, the treatment yields results parallel to those of Gardenfors. (Gardenfors, p. 73.)

2. The Probabilistic Alternative.

To be contrasted with this approach in terms of deductively closed sets of propositions, we may consider a purely probabilistic construal of knowledge: We take a statement as acceptable in our knowledge base when it becomes overwhelmingly probable. This is in accord with the nearly universal agreement that when it comes to empirical matters of fact, there is nothing (or almost nothing) that is certain. Almost any of the things we take for granted "could" turn out to be wrong. Nothing is incorrigible. Not even "observation" statements: without knowing how to handle errors of observation, modern science could hardly get off the ground. Of course, very crude observation statements, e.g., "the sun is shining now," are very unlikely to require correction. (They could be wrong: my "observation" may result from post-hypnotic suggestion, rather than the state of the weather.)

One way of dealing with an approach to knowledge that takes nothing empirical to be incorrigible is to become a thoroughgoing Bayesian: Represent knowledge as a probability function defined over the whole algebra of propositions in the language we are using for knowledge representation. Of course, as Carnap observed (1951), we must suppose that all refinements have been made in the language: we cannot introduce new terms without risking having to change our probability function. Then when experience causes us to shift the probability of some proposition, that change in probability propagates through the algebra in accord with some rule of propagation. (One possibility is "Jeffrey conditionalization.")

This approach to corrigibility has a number of drawbacks. The main one is computational. In language capable of representing some piece of common senses knowledge, or of reasoning about even quite a limited domain, the computational resources needed mount dramatically. The number of possible worlds, describable in even a constrained language, is LARGE. There is also the problem of the source of the original probability measure. Experts? There is the problem of soliciting consistent opinions. Generalize to sets of probability measures? This might be some help, but perhaps not much. There is the problem of updating: No set of probability assessments is likely to be consistent; adjustments will have to be made to achieve conformity with the probability calculus; and one of the items most natural to adjust is the ratio of probabilities $P(A \& B)/P(B)$; but this is just the

important probability of *A* given *B*. And supposing a collection of agents with a common goal, sharing knowledge: how are disagreements concerning probabilities among these agents to be resolved?

These are difficult questions, and while one cannot be certain that plausible answers can't be found, it seems at least worth while to explore an alternative strategy. The alternative that has been explored for some years is that of adopting a purely probabilistic rule of acceptance: In general, "Accept *P* when its probability is high enough." (HEK 1961)

One question rises immediately: how probable is "high enough?" A tentative answer to this question ("It depends on how much is at stake in using the corpus of knowledge in question") has been outlined in (HEK 1984).

A less immediate question arises when we reflect that probability itself — especially evidential probability — depends on evidence. What is probable depends on what we know; and we are proposing that what we know depends on what is probable. Can we have it both ways? In particular, can evidential probability be serve both functions?

We answer yes. It has been proposed (K 1983, K 1984, K 1974) that having fixed on practical certainty, we can introduce evidential certainty as the square root of practical certainty. (This stems from the fact that, using a probabilistic rule of acceptance, the conjunction of a pair of statements that do not appear conjoined in a higher level corpus will appear in a lower level corpus.)

A purely probabilistic rule of acceptance does not yield what Gardenfors has called "belief sets." The set of accepted statements is not closed under deduction, nor — what comes to the same thing in a logic with compactness — is it closed under conjunction. In general, it is not the case that if *A* and *B* are in our corpus of knowledge, their conjunction will also be in it. Of course it does not follow that the conjunction of a pair of statements in our corpus of knowledge will *not* be in it! There may be large conjunctions of statements whose probability is high enough to qualify for acceptance, and every conjunct of such a set of statements will also be in the corpus. In fact, every logical consequence of *each* statement in our body of knowledge will also be in it.

An immediate consequence is that there is an axiomatic representation of our body of knowledge. That is, there is a (presumably finite) set of statements from which the entire contents of our body of practical knowledge follows. This fact has useful consequences when it comes to talking about revisions of our body of knowledge.

The failure to embody deductive closure is not entirely unintuitive. Our confidence in the conclusion of an argument that involves many premises tends to decrease, even though we cannot put our finger on a specific doubtful premise, as the number of premises decreases. There are good intuitive grounds, even, for thinking that the set of statements that I am well justified in accepting is inconsistent; if it is inconsistent, to apply deductive closure to it would be a disaster. One particularly natural example concerns measurement. Suppose the method *M*

yields errors that are distributed approximately normally with a mean of 0 and a variance of .04. Consider a set of applications of that method, from which we infer, in each case, that the length measured lies in the interval $r \pm .8$ (i.e., within four standard deviations of the observed value.) Surely, by any ordinary standard, these results are acceptable. But if we accept a large number n of these results, it will also be overwhelmingly probable that at least one of them is wrong — according to the same distribution. The resulting body of knowledge is inconsistent.

The picture we work with so far is this: There are two sets of sentences we use to represent our bodies of knowledge. One, the practical corpus, contains the other, the evidential corpus, as a part. Everything in our evidential corpus is also in the practical corpus, since an item is a member of the practical corpus if and only if the lower found on its probability (since we are using evidential probability), relative to the evidential corpus, exceeds some fixed probability p.

Statements may come and go, in the practical corpus, according as their probabilities vary with the contents of the evidential corpus. Thus there is no direct problem of revision, expansion, or contraction: all are taken care of by the probabilistic rule of acceptance.

This applies to statistical statements, as well as other statements. So we will have such statistical statements in our practical corpus as "about 95% of birds fly," "less than 2% of penguins fly," etc.

Now how about the corpus of evidential certainties? How do statements get in this corpus? By being probable enough, if we are to have a uniform treatment of acceptance and corrigibility. But we can't (for reasons pointed out in HEK 1961b) just consider simultaneously a sequence of bodies of knowledge. So we must construe a question about the contents of the evidential corpus as shifting context: now we are thinking of a different and higher level as the "evidential" corpus, and what was the evidential corpus as a practical corpus.

3. Probabilistic Inference.

Statistical inference is no problem for evidential probability, but there is no ordinary way that empirical generalizations ("All Crows are Black," "Length is additive under collinear juxtaposition," etc) can be given probabilities. And it is just such items of knowledge that we would like to be able to correct. A related fact is that epistemological probability is defined only relative to a fixed language: the definition is syntactical, and depends on the recursive specification of potential reference classes and potential target classes. How do we handle generalization? And how do we deal with the relativization of probability to a language?

The key notion is that of error. We do not suppose that we have a clear cut distinction between "observational" predicates and "non-observational" predicates. We suppose instead that there is a metalinguistic corpus, parallel to our evidential corpus, that contains a representation of our knowledge concerning observational error. For example, it is there that we store the knowledge that method M for

measuring length yields errors approximately normally distributed with a mean of 0 and a variance of .04.

The details of this construction are to be found in K 1983 T and M. The general idea is that empirical generalizations and theories are construed as features of the language we choose to use. But to each of those possible languages will have going along with it, based on a given stock of actual experience, a corpus of knowledge concerning observational error. Good "observational" predicates are those that can be used with little chance of error; "non-observational" predicates will be those that have significant errors associated with them.

Observational error is generated by the interaction of our experience and a language in the following way: We know that error has occurred when we make a set of judgments that cannot all be true. Thus

What we need, then, is a way of choosing between candidate languages on the basis of the consequent errors associated the languages. In earlier work (T and M 1983, 1990) we approached this question in a very abstract framework, with a view to obtaining treatments of error in both direct and indirect measurement. Here we will adopt the same general standpoint, but examine a variety of replacements of framework assumptions (and expansions and contractions) that are rather more specific.

Our alternative approach has been briefly hinted at in Measurement and Science and Reason. The basic idea is that conflict between a general framework or model, and a set of routines of observation, is reflected in what we take to be the reliability of these routines. Thus if there is a lot of stress between our view of the world and our observational routines, we will be forced to conclude that our observational routines yield a significant amount of error. Given a choice between two frameworks, we choose that that minimizes this error.

This approach does not require either measures of information defined on our language (or languages) or subjective measures of probability. It proceeds in terms of classical statistical inference and evidential probability, and requires only a single index, corresponding to Levi's index of caution q. For present purposes, we will suppose that the observational routines are fixed.

An observational routine is a procedure for adding information to the corpus of knowledge K. Observation and measurement are the prime examples. In direct observation, something happens to you add a (possibly complex) sentence to your corpus of knowledge K. Under no circumstances does observation yield an incorrigible result; error is always possible — as an extreme, hallucination. But under ordinary circumstances, observation does yield knowledge. Indirect observation — observation through a telescope, or a microscope, or a radar screen, or contact lenses, admits of error, but yields knowledge. Similarly, measurement, though it always admits of error, yields knowledge. We measure the voltage, and obtain a value of 3.15, and conclude that as a matter of fact the voltage is between 3.12 and 3.18.

It is enlightening to reflect on measurement. Ordinarily, one supposes that errors of measurement are normally distributed with a mean of about 0, and a variance that is characteristic of the measuring process or instrument. But if these errors are normally distributed, an error of any finite amount has a finite probability. Given our measured value of 3.15 volts, there is a finite probability that the actual voltage is 3000: not at all between 3.12 and 3.18. How did we conclude as knowledge, as a matter of fact, that the voltage was in those limits if it could have been outside them?

We did so on the basis of high probability. That is: there is a number q determined by context in ways that will be considered later, such that if the probability is less than 1 - q that something is so, we just dismiss the possibility. Thus, having made the measurement, having no reason to suspect anything peculiar about it, we just dismiss the possibility that the true voltage is 3000.

There is much to be learned from this simple example. First, we accept limits on the voltage. We do not merely assign a "high" probability to the claim that the voltage lies within those limits. We use this claim as a premise in arguments: if the voltage is less than 3.20, then the the solenoid will not operate; the voltage is between 3.12 and 3.18; therefore the solenoid will not operate. We go on to make further inferences, with the help of more premises: therefor the starter motor will not engage; therefore the engine will not start; therefore ...

In principle, we could avoid acceptance. We could assign probabilities to each of the statements in our cascade of inferences. This is not the way people seem to operate. But it isn't clear how much ice that observation ought to cut. What seems likely is that keeping track of probabilities is just computationally infeasible except for relatively small algebras. In fact this fact might well be the biological reason that people have evolved to argue in logic rather than in probabilities. But that is a matter of speculation. In any event, there is good reason to explore an acceptance model of belief in addition to a purely probabilistic model in which changing degrees of belief migrate over a field of propositions.

Second, the basis on which we accept the limits on the voltage is a straightfoward statistical law: errors of measurement characteristic of the process we used to measure the voltage are distributed according to the distribution D. This also is something we accept; we presumably accept it on the basis of a body of evidence; we presumably accept it because it is overwhelmingly probable. But what is the relation between accepting the law of error and accepting the limits on the voltage?

In principle, they could both be reflected in the same structure. We could have a body of evidence, that would include both the evidence on which we base our statistical law of error, and the evidence comprising the measurement in question, and, relative to this body of evidence we could accept both "The distribution of errors of measurement is D" and "this voltage lies between 3.12 and 3.18 volts." But note that in this case we would not be basing our statement about the voltage on the "known distribution of error." Rather, both the statement about the distribution and the statement about the voltage would be based on a single body

of evidence. Furthermore, it is not easy to see how in this framework we can find a basis for accepting a distribution of error: How do we know when we have made errors? How do we know how big they are? Again computational problems loom.

A simpler structure is obtained by representing our knowledge in two levels: an evidential level and a level of practical certainty. (Kyburg, 1974) At the level of evidence, we accept both the result of the individual measurement and the statistical knowledge reflecting the reliability of the class of measurements to which we take the individual measurement to belong. We must first account for this statistical knowledge.

4 New Observations

There are a number of ways in which new data can impinge on our old body of knowledge. The most common is simply to have new observations added to our body of knowledge. This has an impact on what we believe even when it does not contradict anything we already belief. This impact has two forms. To accept the observation that A is a crow and that A is black entails, in our body of knowledge that A is a bird, since we know that all crows are birds. What is entailed by our background knowledge, and the new observation, becomes part of our background knowledge. (Subject to some caveats we'll get to later: the consequences of long conjunctions of things may not be in our body of knowledge.)

The other form, more interesting in this context, is the impact that the observation has on our general statistical background knowledge. If we have statistical beliefs concerning the frequency with which A's are B's —e.g., that it is between p and q — and we observe an A that is not a B, that should change our body of knowledge, but not very much. If we had earlier accepted our statistical knowledge on the basis of an observation of n A's, of which m were observed to be B's, we now have, as a basis for our statistical knowledge about A's and B's a sample of $n + 1$, of which m are B's. It is clear that our body of knowledge will change relatively gradually as new observations come in: we will not, in this context, find the discontinuities that we observed earlier.

There is also the possibility that our background knowledge, even though statistical, is based on more than observation. For example, my belief that the chances of a birth being the birth of a male is about in [.50,.52] is based on lore obtained from sources that I regard as reliable. To learn that my daughter just gave birth to a boy will not only have little impact on that statistical generalization: it will have *no* impact. But if my source of knowledge were impugned, that would have a large effect. And it is conceivable that I could myself acquire such a large database of sex observations that my own data would impugn the authority on which I had accepted the conventional interval.

This also applies to the sort of statistical knowledge based on physical principles and assumptions. If a die is well balanced, then the velocities and momenta that characterize its trajectory will lead its landing on each side with very nearly equal frequency in the long run, in view of the fact that very small changes in

these momenta will lead to discontinuously different outcomes. If I roll a die and get a '1', my beliefs concerning its statistical characteristics will be unchanged. (Contrary to the Bayesian view, which would demand a tiny change.) If I roll the die a lot, and get a disproportionate frequency of '1's', then at some point I will question my assumptions — in particular, the assumption that the die is well balanced — and replace (not modify) my belief that the long run relative frequency of 1's is 1/6, by a statistical belief determined by my experience. (This will not be a very exact statistical belief, since I may well make this replacement on the basis of a fairly small sample. Thus I might come to believe that the frequency of 1's is in [.5,1.0].)

Thus even in the case of statistical knowledge, augmented by some more instances, there may be discontinuities. We have continuity (and, strictly, even this is not usually continuity in the mathematical sense) only when our evidential knowledge base contains representations of all the data on which our the statistical law in our practical corpus is based, and when, in addition, we obtain additional statistical evidence by a procedure which is evidentially reliable.

Let us consider the other cases, in which our statistical knowledge is modified by the acquisition of new data. Suppose we have in our evidential corpus a statistical law — e.g., that the proportion of *B*'s among *A*'s lies in the interval [p,q]. We observe, with evidential certainty, an *A* that is not a *B*. This has, and should have, no effect on our statistical knowledge. The probability of our evidential statistical knowledge is [1.0,1.0], and thus quite independent of the outcome of a particular trial. (Thus *evidential independence* is not symmetrical! The probability of the outcome of a trial obviously depends on our statistical knowledge.)

We observe a lot of *A*'s, some of which are and some of which aren't, B's When should we take this as evidence bearing on our statistical knowledge? Here is one plausible idea: Suppose that the statistical law is in the evidential corpus whose level is *r*. That is the corpus into which we accept things, provided the chance of error is less than 1-*r*. Suppose that what we have observed is a priori less probable than this. Before the event, we are practically certain that we won't observe what in fact we observed. In itself, this does *not* impugn our generalization: the improbable does happen, and there, before our very eyes, is an instance of it. Besides, as Savage pointed out long ago, whatever happens, described in detail, is extremely improbable. The most pedestrian bridge hand has only a chance of 1 in 10^9 of being dealt.

What leads us to question the fairness of the deal (that is, the appropriateness of the usual statistical law governing bridge hands) when we get 13 spades? But not when we get ♣ 9, 3 ♦ K, J, 2 ♥ A, 4, 3, 2 ♠ 9, 6, 5, 3, even though the chances of getting this hand are just as small as the chances of getting 13 spades? That's a good question. The answer usually given is that there is an alternative explanation for the 13 spades (somebody is cooking the cards), that renders that particular hand more probable than it would be in the course of nature. But this is just to look at a corpus in which we can evaluate the relative likelihood of the ordinary statistical law and the fall of the cards being the result of manipulation. What happens when

we encounter *striking* evidence (not merely 'improbable' evidence) is that we are led to alter our level of evidential certainty to accommodate the *probability* of what we ordinarily take as evidence.

This move is always available to us. But it is one we will only make when we are motivated to make it. In accordance with the analysis offered in (Theory and Decision), we will be motivated to raise our acceptance and evidential level when there is more at stake. That's not hard to see in this context: If we are playing high stakes bridge in a shady gameroom in Reno, we are likely to consider the possibility of cheating more seriously than if we are playing a friendly game of bridge with our in-laws.

Let us look at an alternative circumstance under which we might be led to suspend acceptance of a statistical law. Suppose we have a die that seems perfectly symmetrical, but that turns up five 40 % of the time on a thousand throws. We accept, we assign probability 1.0 to, the proposition that the die yields five about 17% of the time. But here we have an observation that has a probability of essentially 0, given the truth of our assumption. Such observations *do* occur (remember the ordinary bridge-hand), but *if* there is an alternative account according to which the observation is not so improbable, perhaps it is worth our effort to escalate the level of our evidential corpus and examine the probabilities of the alternatives.

What are the alternatives to be examined in this case? There is the possibility that the analysis of the behavior of dice in terms of varying outcomes with varying momenta is wrong: Note that the probability of getting 40% fives, on the usual hypothesis is no less than the probability of getting any sequence of a thousand outcomes. It is *not* the improbability of the observation that leads us to a new possibility. It is the fact that we have alternative statistical laws in mind that render the result more probable. We would not (on the basis of the evidence described) conjecture that the results of the tosses were not independent and identically distributed. We would conjecture that the die was not symmetrical, and that the outcomes of its tosses were multinomial with a parameter close to 0.4 for five.

Of course as soon as we reject the reasonable presumption that the die is fair, we are in a position to start using the evidence we have concerning its outcomes to confirm at the level of evidential certainty a statistical law characterizing them. We find, once more, discontinuity: We do not become suspicious of the die on the first few tosses: four fives on ten tosses is perfectly understandable, if a bit unusual. We do not gradually modify the statistical law that we take to govern the outcomes of the die. At no point, in the scenario described, do we reject the statistical law according to which the outcomes of the die are iid. But at some point we flatly reject the assumption that the die is a standard well-balanced die, and begin to use our data for an inference about its approximate true multinomial distribution.

5. Conflicting observations.

It is useful here to make a distinction between 'observation reports' — what is said to have been observed, and 'observation statements' — what is alleged in the

report to have been observed. Observation reports cannot really conflict. If I report the weight of body *W* on one weighing as 23.654 grams, and on another weighing as 23.655, there need be nothing wrong with my observations, although the observation statements, "*W* weighs 23.654 grams," and "*W* weighs 23.655 grams" are inconsistent. This is why the natural and appropriate observation *statement* is rather, "*W* weighs 23.65 ± .02 grams." Note that this statement is not certain: It is acceptable, because the chance of error is negligible, not because it is impossible. On the usual treatment of errors, they are treated as normally distributed, and an error of any magnitude is *possible.*

We treat it as evidence, however. We take it to be a statement that we can use in designing machinery, in engineering, in prediction, etc. It is not a statement to which we merely assign a high probability.

But it is corrigible. We may weigh *W* twice again, and conclude (with the same degree of justification as we had before) that it weighs 23.60± .02 grams. The two statements are strictly incompatible.

There are various possibilities. First, we may suppose that we simply have made somewhat unusual errors of measurement. If it is evidentially certain that *W* weighs between 23.63 grams and 23.67 grams, then *W* cannot weigh as little as 23.62 grams. But if *W* can't change weight, the discrepancy *must* be due to errors of measurement. If this is the case, then there are two impacts of our conflicting observations: The observations should be combined; and the discrepancy between the two sets of measurements should be taken as evidence concerning the distribution of errors of measurement for the measuring device(s) involved.

Merely combining the measurements would yield 23.62 ± .015, if we assume that all four measurements are simply taken from the same normal population of measurements. But the discrepancy might suggest that we should regard the measurements as coming from two distinct populations, or as coming from a population with a larger variance than we had thought.

In general, the conflict among observation reports must be taken as evidence concerning the reliability of the observer, or of the apparatus, of both. We will find that this is true also in the case of more basic conflicts.

6. Conflict between Observation and an Accepted Framework.

This is the most interesting sort of conflict. It is the sort that is most likely to arise, since we often make relatively local assumptions that we take for granted, act on the basis of, until and unless they lead us into difficulty. Good judgment consists in knowing when to abandon an assumption. But can good judgment be codified, reduced to mechanical rules? In some respects, we will argue, it can.

The simple-minded view of belief change is this: You have a generalization that you have taken for granted that leads you to infer that observational circumstances C will be followed by or accompanied by observational outcome O. You observe C. You observe some contrary of O. You reject your assumption.

But things are almost never this simple. Even when (rarely the case) a qualitative generalization is understood to be strict, to admit of no exceptions, there are alternatives to rejecting the generalization in the face of apparently conflicting observation. We may take the alleged observations to have been in error. Illusion, hallucination, are always available to explain away apparent refutations. And this is not irrational. In fact it has been argued that this is the source of our knowledge of the qualitative errors of observation. The identification of an object of observation as belonging to a given *kind* is subject to error. The frequency of such errors is given (as suggested in Science and Reason) by two principles: One is the conservation principle, which suggests that we should not attribute more error to our observations than we are obliged to by the model of the world we accept. (Notice how this is in almost direct contrast to the proposal that one should accept a model of the world only if it does not contradict any of one's observations.) The other principle guiding our assessments of error is the distribution principle, which says that, *given* the satisfaction of the conservation principle, we should distribute the errors we are obliged to attribute to our observations as evenly as possible among the *kinds* of errors we might have made.

Thus if our model of the world assumes (presupposes) that all crows are black, and we have some observations of blue crows, we would assume that those observations contain errors. And further that the errors are (other things being equal) are distributed equally between judgments of blueness and judgments of crowness. The metalinguistic fact that we must assume that we have made these errors of observation provides evidence about the *reliability* with which blueness and crowness can be identified.

Naturally things are not this simple, since in the real world "is blue" is a vague term, and furthermore a term that enters into a great many rough generalizations; and "is a crow" is a technical term of ornithology, involving a complex set of necessary and sufficient conditions that are tied to a great many other properties, some of which are observational and some of which are not.

Now of course there are circumstances under which observations of blue crows would lead to the rejection of the frame assumption that all crows are black. This is exactly the kind of thing we are looking for: when are things so anomalous, given our assumptions and beliefs, that we should profoundly alter those assumptions. So let us look at two new crow stories.

The simple story continues to make use of ordinary observation. Suppose that many people report seeing many apparently blue apparent crows. Their reports are wrong, of course, given our assumptions. But two facts about error are entailed by the prevalence of people seeing 'blue crows'. First, the reliability with which people can identify the color blue decreases, in general. The same is true of the reliability with which ordinary observation can identify a crow. The long run inferred frequency with which observation reports of the form "x is blue" are in error, after our experience with the blue crow observations, is higher than it was

before. Similarly for the long run inferred frequency with which observation reports of the form "*x* is a crow" are in error.

But this is a small difference (we suppose) and in fact is not the relevant one. Observations are made in context, and we can find a context in which a subset of the class of "x is blue" observations (when they occur in conjunction with an"x is a crow observation") in which the frequency of error is very high. (Perhaps 50%, since any conjunctive blue crow observation, given our assumption, must contain an error in at least one conjunct.) This is such a high rate of error that we do *not* have observational grounds for accepting "*x* is blue" or "*x* is a crow" in this context. When John reports to me that there is a blue crow on the fence, I not only must reject the observation statement corresponding to his report, but even the observation statement that there is something blue on the fence and the observation statement that there is a crow on the fence. This is a serious loss of communication as well as an impoverishment of our language.

We lose something by abandoning the generalization that all crows are black; but we lose more by abandoning the reliability of observation terms in a certain context.

It is not always true that we will abandon the generalization in favor of the universal reliability of observation. We abandon the reliability of perceptual judgment of straightness when it comes to sticks half in and half out of the water, because there is such a rich matrix of generalizations concerning sticks (!) that we would lose more by abandoning those generalizations than by supposing that our perceptions of sticks in water are unreliable. (Note that this is true even without the knowledge of why the stick appears bent; we did not require a theory of refraction to know that putting a stick in the water didn't bend it.)

The second crow story is more complicated and also more realistic. We would not in general require that a lot of people reported a lot of blue crows before we abandoned the assumption that all crows are black. One, or one or two, good crows, observed in careful scientific detail, would do it. Black and blue are hard to distinguish, sometimes, and most of us are not ornithologists. But we will not suppose that a trained ornithologist will make a mistake about even a single specimen. What is the difference?

The difference can be explained in terms of probability. Here is R. A. Fisher's explanation. (scientific inference) G entails something very improbable, C, that in fact happens. The improbable doesn't happen. Therefore G is likely to be false. Both scenarios fit this description: Ordinary observation is unlikely to be wrong about such things as blueness and crows, but not *very* unlikely. On the other hand, a lot of ordinary judgments are *very* unlikely to be uniformly wrong. Professional scientific observation is, from the outset, *very* unlikely to be wrong.

As we have already seen, this is an oversimplification, since what actually happens is always very unlikely (the bridge hands). What is left out of account is the existence of an alternative framework, assumption, presupposition, according to which what we have observed is not so unlikely. In the case of the blue crows, the

alternative assumption is quite simple: abandon the general assumption that all crows are black, and replace it with an approximate statistical generalization concerning the frequencies of blue and black crows. This results in a significant loss of information in one regard, especially since in the beginning there will be little statistical evidence to base our generalization on. But it would be worse to stick to the universal generalization in the fact of a most improbable collection of observational errors.

To contrast: in the case of the stick that appears to be bent in water, we have no alternative assumption that doesn't entail severe damage to our body of knowledge. (Sticks have a lot of properties — rigidity, relatively constant bending moments, etc. — that are inconsistent with the accuracy of our perceptions of sticks in water.) So even at the cost of a whole class of unreliable perceptions, it is better to continue to believe that putting a stick in water doesn't bend it.

8. Quantities conflicting with formulas.

Suppose in general that we assume the quantitative law, $y = f(x,z)$ in our body of knowledge. Then we observe a series of measurements of the quantities X, Y, and Z. No set of measurements can contradict the law in question, since any measurement is subject to error, and indeed, on the usual theories of measurement error, subject to error that can *possibly* be arbitrarily great. But of course large discrepancies, relative to a body of knowledge that contains the law in question, are extremely improbable.

The same general approach makes sense: The very improbable happens all the time (the particular set of measurements we make are improbable even if our assumed law *is* true), but if there is an alternative that renders the improbable not so improbable, the observations support that alternative. To put a quantitative measure on this is not trivial. One way, in terms of the framework we have already talked about, is the following: Anomalous observations can have two effects: they can provide new data concerning the errors of observation of a certain sort, or they can be taken at face value, and thus provide grounds for the rejection of general formulas.

7. **Fundamental Assumptions.**

Before going on to consider the grounds on which one would choose to give up an assumption in favor of attributing errors to one's observations, it is worth looking at one more extreme cases. This is that of measurement, and has been discussed more fully in (Theory and Measurement). We suppose that length is additive: that the length of the collinear juxtaposition of two bodies is the sum of their lengths. Our measurements, of course, do not support this supposition; less dramatically: we can maintain the additivity of length only by attributing error to almost all our measurements.

Is this the alternative? To suppose that we can measure accurately, but that length is not additive, on the one hand, or, on the other, to suppose that length is

additive, but that all our measurements are infected with error? Put this way it seems odd that one would ever opt for the second alternative. But we do.

Here is a possible explanation. The errors of measurement we need to introduce are very rarely large. They therefore do not deprive us of much useful knowledge. But the additivity of length is an enormously powerful predictive device. Knowing the length of two rigid bodies, we know, without even measuring, the *approximate* length of their collinear juxtaposition.

The choice between attributing error to observations and maintaining a generalization, as opposed to taking observations to be accurate and to refute the generalization, lies in the predictive observational content of the whole body of knowledge involved.

How do we measure predictive observational content? That seems to admit of no simple and general answer. In the case at hand, though, the measurement of length is so pervasive, and the predictions we get from construing length as additive are so widespread, that there can be no question about the choice. What we would like to achieve is a principle or set of principles that apply to less obvious cases.

8. **Choosing between an assumption and errors or between assumptions.**

Suppose we consider two bodies of knowledge, one that embodies among its evidential certainties (among other things) the assumption A, the other of which does not. We make a set of observations (add to our evidential certainties a set of observation reports). We have in our background knowledge statistical information about errors in observations of this sort. Given the assumption A, the observation reports must be taken to embody unusually (improbably) large errors. These errors are not without observational consequences. They render observational predictions less dependable, since the correspondence between what is predicted and what probably going to be observed is only approximate, and reflects our knowledge of errors of observation.

Here are the two cases: Keep assumption A, and suppose that errors of observation in the circumstances are large. That large errors of observation are encountered in this situation provides evidence that the errors of observation in this situation are unusually large. (We base our knowledge of the frequency and magnitude of errors — their distribution — on the sample we have, and these errors form part of that sample.) That means that our predictions are less precise, and thus less useful. This applies not only to predictions made in accordance with assumption A, but predictions of the same sort made under the same circumstances whether or not they involve assumption A. This may be worth it: half a loaf is better than none.

Second, give up assumption A. Interpret the results of your observations as refuting A. Now we need not impugn our observations either in general or in the particular circumstances at hand. Our observations are as accurate as they ever were. Everything else, we may assume, remains unchanged in our body of knowledge, and thus all we lose are the predictions based on assumption A.

How do we weight the advantages of one choice or the other? In order to have an actual measure that will yield an answer in these cases, we must focus on a class of predictive statements — that is, a class of statements that is of interest to us in the circumstances at hand. It is in this class that the predictions of the two cases are to be drawn. Let this class be C. We also need a measure of the precision of the predictions: thus if a prediction has the form "Bird B is Blue," the amount of content of that prediction must reflect the chance of an error in the observation that would test that prediction. If we can't accurately tell blue things, there is less content to the prediction that something is blue. If the prediction has the form, "Object O will be observed at an angle between a – d and a + d," then its content will reflect the distribution of errors of observation of angle in the circumstances under consideration.

The class C of predictive statements about which we are concerned should be finite. It can be large, but we want to ensure that ratios are well defined in it. How do we characterize this set of statements? I don't know, but it clearly should be context dependent. Next we need measures of accuracy.

With regard to categorical statements ("There's a crow," "that's blue") we get two cases: there is no prediction (clearly no help at all) or there is a prediction that reflects a certain error rate, or pair of error rates, in using the term being predicted. We refer to a pair of error rates, since there is both the chance of failing to identify an instance of the predicated predicate or relation, and the chance of falsely attributing to an object the presence of the predicate or relation. One natural approach would be to regard each kind of error as being equally important. But this may not be appropriate. In a given kind of context, one of these errors may be much more important than another. That difference of importance can be reflected in the cost of errors of the two kinds. It cannot be given a priori.

With regard to quantitative statements ("The widget will be observed at angle alpha," "An increase of weight will be observed") there is a standard conventional measure of the error: namely, the square of the difference between the predicted value and the observed value (the value indicated in the observational report). But again, this should perhaps not be taken as universal. It may be that in a given kind of context, errors in one direction are much more important than errors in another direction.

What we need is only (i) a (finite) set of sentences that include all those that may be of predictive interest in a given context, and (ii) a measure of how important errors of various kinds are. We get the frequencies of these errors from our background knowledge of the observation reports we have had, together with assumptions of our body of knowledge. When we change the assumptions (or eliminate one) we change the statistical representation of these errors that we have reason to accept. If, for example, we eliminate an assumption,we can replace a number of predictions (those that stemmed from that assumption) by no predictions.

9. Summary.

Global approaches to replacing one theory by another require relatively universal conventions: an ordering of all the sets of sentences in a formal language, for example. Approaches to eschewing acceptance, and therefore replacement, such as proposed by "Bayesian probabilists" tend to be impractical for simpler reasons: too much computation is devoted to issues that are at best peripheral to the questions at hand ("Should we assume that instrument I is operating correctly?").

We have proposed instead an approach characterized by a set of sentences (sentences that could, in principle, be construed as predictive observational sentences in the sense characterized above), and also by a measure of informational value determined by a distribution of errors for these sentences. Given a pair (C,m) consisting of a set of sentences and a measure of the importance of errors, then the relative value, in the face of a given body of observational reports, together with a body knowledge, of two assumptions, or of one assumption as opposed to none, is determined. It is determined by machinery of probability that we already have in hand.

There is, of course, the problem of determining the pair (C,m) to fit a given context. We have not yet dealt with this problem. We observe only that it is a far less overwhelming problem than that of determining informational content of all the sentences of a language (Levi) or of associating with each sentence of the language an ordinal number (Spohn). It can be done for a specific class of circumstances when certain kinds of predictions or anticipations are the kinds at issue. When the "assumptions" about which we are talking are relatively limited in scope ("Instrument 47 is working correctly"), it is not at all unreasonable to suppose that in fact we can isolate such a useful set of sentences. The question of deriving such a set of sentences from our concerns in a given context, and the question of deriving the importance of various kinds of error from the utilities of the outcomes possible in a given context, are questions that must be reserved for another time.

Henry E. Kyburg, Jr.
University of Rochester

REPORT DOCUMENTATION PAGE

Form Approved
OPM No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1989	Unknown

4. TITLE AND SUBTITLE

Giving Up Certainites

5. FUNDING NUMBERS

DAAB10-86-C-0567

6. AUTHOR(S)

Henry E. Kyburg

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Rochester
Department of Philosophy
Rochester, NY 14627

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army CECOM Signals Warfare Directorate
Vint Hill Farms Station
Warrenton, VA 22186-5100

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

92-TRF-0015

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Statement A; Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

People have worried for many years - centuries - about how you perform large changes in your body of beliefs. How does the new evidence lead you to replace a geocentric system of planetary motion by a heliocentric system? How do we decide to abandon the principle of the conservation of mass? The general approach that we will try to defend here is that an assumption, presupposition, framework principle, will be rejected or altered when a large enough number of improbabilities must be accepted on be basis of our experience. If I think that all swans are white, and a student claims to have a counterexample, I will assume that he has made some observational error. I will reject his result, and continue to accept the generalization. When a lot of people claim to have seen counterexamples, I will come around: to continue to accept the generalization would require me to accept too many improbabilities. This is a discontinuous probable, while certain reports become more probable. We cannot accept the generalization and even one of the observation reports: that would be a simple inconsistency.

14. SUBJECT TERMS

Artificial Intelligence, Data Fusion, Certainities

15. NUMBER OF PAGES

20

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL