Potential Surprises

Frank Hampel
ETH Zurich, Switzerland
hampel@stat.math.ethz.ch

Abstract

After a brief historical overview over various approaches to the foundations of statistics, the very general, simple and basic concept of (potential) surprises is introduced, which may be subjective or objective and goes beyond previous approaches by I.J. Good and by the author. The surprises are conditional on the background knowledge or belief of the person experiencing it; the updating of the so-called background, and the merging or, if not possible, the contrasting of different backgrounds by two or more persons (otherwise they talk past each other) are very important operations in practice. A number of examples from real life, in complement to two previous, more qualitative papers, are given.

Keywords. Foundations of statistics, historical concepts, (potential) surprises, background knowledge or belief, combining of backgrounds, updating of backgrounds, merging or contrasting of backgrounds, practical application of mathematical models, real life examples.

1 Introduction and overview

Over the centuries, there have been various different approaches towards the fundamental concepts of statistics.

One line of thought focusses on observed and hypothetical frequencies of "random" events, especially – usually under some symmetry assumptions – the expectations for games of chance. (I shall leave aside the various philosophical meanings of "randomness".) After some previous isolated attempts, this led to the work by Pascal ([33], cf. also [18, Ch. 8], and also [3, Part 4, Ch. XVI, p. 387-388]) and Fermat (starting 1654), Huygens, de Moivre, Laplace, and later the frequentist theories by von Mises, Neyman-Pearson (cf., e.g., [31]), and Wald, among others. A basic tool was the law of large numbers [5, Part 4.5] and its later refinements, which in practice allowed to approximately equate the observed percentage of successes in a "long" sequence (whatever that means, cf. [22, Part 5]) of random experiments with their theoretical probability.

Besides a lot of mathematical work building upon the basic assumptions, there is still a chance for new ideas about this line of foundations, as shown by Cattaneo's [8, 9] improvement of Wald's minimax principle.

A very different approach to the foundations of stochastics, which apparently has not found much attention, goes back to Jacob Bernoulli [5, Part 4]. Besides continuing the work of Huygens (and discovering the law of large numbers, his "theorema aurea", as an auxiliary tool for something very different), he tried to develop a quantitative counterpart to the (then very famous) dichotomic "Logique" by Arnauld and Nicole [3]. His aim was to measure the degree of "probability" in the old, qualitative sense of this word (cf. [7, Ch. 2.8, 5.2, 5.3], also briefly [30, Ch. 2, espec. first paragraph, and Ch. 4.3]). Apparently by a misunderstanding of the eloges at the death of Bernoulli in the year 1705, the term "probability" was then used also for games of chance (cf. [7, Ch. 6.4.2]). (It is interesting to observe that both the terms "probability" and "statistics" (cf. [35, pp. 2f and 8f]) originally had a very different meaning.) Bernoulli proposed in a normative way in words (not in formulas) 9 "axioms" or basic (self-evident) properties which the new probability ought to obey and which would exclude, for example, both the Bayesian and the Neyman-Pearson theory and would not even obey in general the rule of additivity (cf. [7, Ch. 5.3.2]). But he still counted cases, as is done in games of chance. Perhaps he hoped to be able to derive "objective" results. Altogether, his approach (left incomplete because of his death) is a bold, fascinating and singular but perhaps shaky edifice; it seems not clear to me whether it can be worked out to a fully functioning system.

Another approach has become highly influential, namely the approach by Bayes [4]. Contrary to common belief, Bayes was basically a frequentist; only in his famous Scholium did he leave some open questions between the lines (which may be even the reason for not publishing his paper during his lifetime) which led to the later "Bayesian" interpretation (cf. [23, Ch. 1.3]). But cf. also the critical remarks by Boole [6].

Variants of the "Bayesian" approach were used by Laplace, Jeffreys (both normative or "logical" in different ways) and de Finetti (subjectivistic). Especially de Finetti [11] was a very sharp, radical, philosophically deep and fascinating thinker, building a pure and clean theory (although he did also applications). But if his theory is taken literally, I find it in its last consequence solipsistic, without any relation to anything like a "real world" – which for him does not exist - or to any fellow scientist. (I only know from L.J. Savage, one of his main pupils, that Savage was sometimes pragmatic in applications; moreover, his (the latter's) pupil D. Ellsberg showed with his paradox that some basic assumptions of Neo-Bayesians do not work in practice.)

All Bayesians (including "logical" or "objective" Bayesians) consider only epistemic probabilities (referring to our knowledge or belief about Nature, not about the (principally unknown) state of Nature itself. (Empirical Bayesians use frequentist methods.) On the other hand, all (traditional) frequentists consider only (usually unknown) aleatory probabilities in Nature, without any reference to what we know or may know (the few cases with known or strongly believed probability models excepted). Perhaps the first one to build a formal bridge between aleatory and epistemic probabilities was R.A. Fisher [15] with his fiducial argument. Unfortunately, he later made a mistake in its interpretation, but this mistake can be corrected, and Fisher's (corrected) fiducial probabilities just turn out to be a very special case of a general theory, using upper and lower probabilities [21, 23, 24, 26, 27].

It seems still unbelievable to many mathematical statisticians that one can derive known epistemic probabilities from unknown aleatory probabilities; but this is correctly done by most intelligent users of statistics who have not given up their own intuition in favor of either the Neyman-Pearson or the Bayesian theory (which, to be sure, are correct as far as they go, but in my eyes do not cover all the needs of good applied statistics, cf. [21, Ch. 4, p. 130], [23, 1.3], [24, Ch. 1.1]). And it has been done long ago, also at the early time of Fisher, cf., e.g., Student [38] or Pearson & Wishart [34]. Even though Fisher seemed only intuitively and not rationally clear about it, his concept of confidence intervals was clearly epistemic (and hence allowing a correct "aposteriori" interpretation), while that of Neyman was clearly aleatory, explaining Fisher's original doubt and later his conviction that despite all superficial formal similarities the two concepts are indeed different, referring to two different probability spaces.

In recent times, there are a number of approaches to statistics in a very broad sense using something like upper and lower probabilities, instead of strictly additive probabilities, as is only too well known at ISIPTA conferences (cf., e.g., [14, 37, 39, 40]; cf. also the ISIPTA conferences). Also some other statisticians, although claiming to be Bayesians, occasionally or inconspicuously use upper and lower probabilities, notably Dempster [12, 13] and Good (cf., e.g., [17]). There are a number of different concepts defined and many results developed. Also one of my lines of work [21, 23, 24] which centrally uses bets (like the Bayesians), but introduces also one-sided bets (thus leading beyond Bayesians) and uses also upper and lower fiducial probabilities, bridging the gap between aleatory and epistemic probabilities, belongs to this body of research.

A main goal of the present paper is to present several new concepts, with the help of practical examples, which ought to be able to describe the inference process on a higher level (cf. also [28, 29]). Although in my opinion inductive reasoning will be done mostly in a qualitative or semi-quantitative framework (using a discrete ordinal scale) as in the previous papers, an attempt is made specifically in this paper to allow also the introduction of a quantitative theory, still leaving a lot of freedom for the precise choices in detail.

There is a concept which looks so simple and at the same time so basic that it seems surprising that it is not more popular in statistical theories: the concept of potential surprises, or of surprises, for short. It can be used as a generic, rather encompassing term; in special situations, it can also be defined as, for example, minus the logarithm of a probability, then giving it a quantitative interpretation. This interpretation is of course inherent in information theory, though it is not normally given a special name. I.J. Good ([16], but not [17]), in the spirit of pure mathematics, has defined a whole mathematical class of surprises. A related definition of surprise is independently given by Hampel [21, Ch. 5]; it turned out that it differs from what Good (orally) considered his most important special case just by an additive constant. But surprises in my present theory can be given any subjective (numerical) interpretation (as is the case with beliefs, subjective probabilities, etc.). This concept, which has been neglected so far, is qualitatively (and semi-quantitatively) investigated in Hampel [28, 29], with a number of practical examples. Both papers are in close connection with the present paper. It is my belief that (like elsewhere in stochastics) the precise numbers don't matter so much as the more qualitative features. (This is shown in the examples of the two previous papers.)

However, some people may want a general quantitative theory, and for this purpose the present paper is written. Yet, to avoid misunderstandings, this paper is basically philosophical and is derived from practical experience of everyday life (including experience in science). It is not derived from some system of axioms. Personally (and perhaps in an oldfashioned way), I do not start with axioms (not even intuitive looking ones as did Jacob Bernoulli), but rather I think axioms should be the crown at the end of the development of a body of knowledge. Later, there were historical reasons for the Bourbaki style in pure mathematics (trying to derive everything quickly on the highest and most abstract level); but I find this even dangerous, as the connections with the intuitive sources, including the nonmathematical sources, easily tend to get lost. As I try to derive all concepts from practical experience, and as this paper is work in progress (with some open questions, e.g. at the end of Ch. 4.2), I shall not try to present an axiomatic development of surprise. (It may even be argued that the problems Jacob Bernoulli had with his approach - see above - may partly be due to premature axiomatization - even though, or perhaps because, he partly relied on the Logique of Port-Royal, cf. [7, Ch. 5.3.2 and 5.2.2.4.].)

This paper contains several examples for the use of the new concepts; for more examples, see the other two papers [28, 29]. One example could even be continued: while the Arctic Warbler (Phylloscopus borealis) in Poland and South China had exactly the same song [29, Ch. 6.5], to my big surprise the same species in Japan had a very different song. It turned out that in Japan breeds a different subspecies (Ph. b. xanthodryas), and it is presently being investigated whether it ought to be separated as a new species from the nominate form Ph. b. borealis.

As stressed already in the previous two papers, the surprises are conditional on the assumed background belief or available background knowledge (both in their intuitive senses), both formally called background for short; and updating of the background, when new information becomes available, is a very important part of the inference or learning process.

The structure of the background is described more fully in [29] and the corresponding poster. Briefly, it consists of all our knowledge, beliefs, conditional or hypothetical beliefs, etc.; but more importantly, it exists in layers, and normally we use only the uppermost layer, containing our most plausible (or likely, or "normal") world view; only when we get a (complete or almost) contradiction with it by some new information (an infinite or close to infinite surprise), we abandon the uppermost layer and fully switch to the next one [29, Sec. 4]. This is (normally) a qualitative change of the background, not just a belief revision (cf., e.g., the article on Belief revision in Wikipedia (05/02/2011)) which slightly modifies the old belief system by means of some logical operations, or information fusion (information integration) or such an operation. It is not a deductive operation, but an inductive jump (cf. the examples); the old theory is false and not just modified, but replaced by something new, created by inductive thinking from the deeper, more hidden layers of our background.

(It might be argued that by enough logical operations one can change the background also to something qualitatively new; but this is not what I experience in the real world examples that came to my mind. I noted already [29, end of Sec. 1] that in the about 20 pages of [1] I could not find a single real life example, while they abound in my papers. To be sure, there is a place for deductive-logical operations; but I suggest that the creative inductive thinking process which generates genuine new knowledge has been badly neglected in research.)

Another important part of the inference process is the merging, or, if this is not possible, the contrasting of the backgrounds of two or more different persons (cf. Ch. 2.3).

As already briefly mentioned above, surprises in my approach are completely compatible with belief theory, Bayesian theory, and so on. They may be seen as a kind of superstructure over the old theories. As long as the surprises (in whatever reasonable way they are measured) are in an intermediate or low range, nothing essential changes. But if they are equal or close to infinity, then the background has to be changed. –

A word or two on terminology: It seems there are too few words in our language for all the different concepts that have been defined. The editors kindly drew my attention to [36] who used not only the term "surprise", but even "potential surprise" (loc. cit., Part II, espec. Ch. IX). There is much overlap and in part(!) a very similar intuition; moreover, style and basic philosophical attitude are quite similar. But I am mainly interested in inference, and Shackle in decisions, especially in economics; his formal definitions are different from mine (for example, by always demanding also a surprise of zero in any disjunction); but most importantly, I could not find the change of background in case of (as he calls it) maximum surprise, which is so central in my approach. He also seems to avoid "...or any other..." which for me opens the door to the radical change of background. Some of his argumentation (against traditional probability theory) now may seem outdated, especially at an ISIPTA conference, and he also has run into problems with his attempt at an axiomatization (cf. above); but overall I find his thinking and arguing quite inspiring, although there is only partly an overlap in our approaches. (At least, the use of the same term does still seem bearable, as long as one is conscious of the differences.)

Another author who introduced the term "surprise" is Neumaier [32]. Again, he just tries to modify the old background in view of contradictions, not abandoning it, by finding an optimal compromise (with an "army of computer slaves" in the fictive story of the king on p. 22), minimizing the total surprise. (This may be appropriate if the surprise is so moderate that the background must only be slightly modified, not abandoned entirely.) And again, Neumaier finds much basic intuition in common with Shackle, but many formal differences. –

Our paper starts with basic definitions, properties and examples, which are not only mathematical since the application of the theory needs also close connections with the nonmathematical world (cf. [22]). Then the case of two or more different background assumptions is discussed, the connection with cautious surprises and successful bets is explained, and the problem of two (or more) persons with different backgrounds is brought into view. The updating of the background information is shown with a complex example, and another complex example asks among other points what to do if an event is totally unexpected. A practical example on how to concentrate incomplete knowledge in a fairly effective way concludes the paper.

2 Basic concepts

The following subsections introduce some basic definitions, properties and examples.

2.1 Basic set-up

Consider one person, say, Ted, with his background knowledge and collection of beliefs B, and a class of uncertain (future or unknown) events E which are of interest to Ted.

Let A be an event in E, and define the nonnegative number $s = s(A|B)$ to be the surprise of Ted, given his background B, when A turns out to be true; with $s = 0$ meaning no surprise at all; $s = \infty$ means Ted considers A impossible; and s "close to ∞" means Ted considers A "practically impossible".

More precisely, we have to distinguish
(i) the hypothetical surprise of Ted when he imagines that A shall happen or (unknown to him) has happened (the potential surprise in the strict sense, cf. also [36]);
(ii) the reaction of Ted when (perhaps in the future) he is reliably told that A has happened (or is for sure going to happen); and
(iii) Ted's reaction when he observes A himself.
(There is not much difference between (ii) and (iii) except the additional reliability by observing A oneself; on the other hand, Ted can also err himself.) Situation (i) requires that Ted thinks of the possibility that A may happen.

There may be possible events which Ted does not even think of. In such a case, Ted's surprise under (ii) or (iii) may still be small if he notes he just has forgotten a rather likely possibility; the surprise shall be very large, if on hindsight he considers the event A possible though highly unlikely; the surprise may be even infinite if A is not compatible with the assumed background B. In this case, Ted has to change his background belief B (one of the most fruitful sources of qualitatively new knowledge), or else he has to change his interpretation of the observation A (e.g., by finding an error in the observation).

In general, we shall change the background, going to the next layer (see above), not only when s is infinite, but also when s is "close to infinity". This is in analogy with what is also called Cournot's principle: that in the applications of probability theory, we consider an event with probability "close to one" as "practically certain" or (formerly) "morally certain". The boundary may depend on circumstances; Bernoulli gives as an example 999/1000 [7, p. 230], while Cournot [10, Ch. IV, 48] requires the difference to one to be "infinitesimally small" for an event to be "physically certain". No matter in what way the surprise is defined, I find the change of background as described the most important application of the concept of surprise.

(A logician might ask what is Ted's surprise by A if he has an "empty" background, e.g. if he wakes up from a coma and has lost all memory and all thinking ability. Then all his surprises are zero, because everything is fully possible. As soon as Ted starts thinking again, one has to be very careful in sorting out what he is able to think and learn again.)

2.2 Some basic properties of surprises

As mentioned above, a surprise s is a real number between zero and infinity, depending on the background knowledge or belief B of a person (here called Ted) and on the (perhaps fictive) occurrence of an event A.

It may be fully subjective, or it may be determined by objective circumstances, yielding an intersubjectively determined number (i.e., the same one for every person with the same background B). In either case, it is an epistemic quantity, that is, it refers to the knowledge or belief of a person, and not to some "objective" property of Nature (unless the two happen to coincide).

Example 1: Let F be a probability space with a known probability $P = P(A)$ for every (measurable) event A in F. Then we may define $s(A|B = F) = -\log P(A)$. In this case, s is a very natural "objective" measure for our surprise in case A happens. Some mathematical properties of s follow in this case, for example, its wellknown additivity. In particular, the expected value of s may be termed the entropy of F. And this entropy may be called the minimum possible average surprise of Ted. If Ted entertains another surprise function s', his average surprise, averaged over all possible events A with their probabilities, will be at least as large.

2.3 Two background assumptions

Now consider the situation that Ted entertains two different background belief systems B and C, perhaps being in doubt which one he should adopt. This may be the case in a learning situation, or in a conflict between different beliefs. If he would be not surprised if either told reliably that B is true, or else that C is true, his (minimum) surprise when observing A is $s(A|B \text{ or } C) = \min(s(A|B), s(A|C))$.

A more refined and more realistic situation is that Ted has different ("apriori") surprises $b(B)$ and $c(C)$ if told that B or C, resp., is true. (The functions or numbers b and c measure the surprise if Ted is told that a specific belief system is true. They may be different numbers, therefore the change of notation from s. In the following we require an additivity property of surprises, as in Example 1.) We call the three surprises s, b, and c unrelated if no occurrence of A or B or C (or a subset of these) affects any other surprise. Then Ted's minimum surprise $s(A|(B \text{ with } b) \text{ or } (C \text{ with } c)) = \min(s(A|B) + b(B), s(A|C)+c(C))$. Naturally, this can also be done with more than two beliefs (cf. 2.4).

The observation A in turn influences Ted's ("aposteriori") surprises about B and C, given A: $b(B|A) = b(B) + s(A|B)$, and correspondingly for $c(C|A)$. (The notation is a bit stretched, as A is not a belief system, but the meaning should be clear.) Naturally, this is close to Bayes' theorem, except that we do not introduce and do not need the renormalization.

The main purpose of computing $b(B|A)$ and comparing it with $c(C|A)$ is that if $b(B|A)$ is infinite, B cannot be used anymore as a background belief (except in case of an error in A); but also if $b(B|A)$ is "much" larger than $c(C|A)$, B is "practically impossible". This is in accordance with common sense thinking (except in case of a very strong prejudice in favor of B which, however, would also imply a very small $b(B)$), but it is at variance with the usual procedure in Bayes theory, belief function theory and similar approaches, where even tiny probabilities or beliefs are being renormalized (as long as they are not exactly zero).

If a model assumption or another basic assumption B is clearly shown by the data to be wrong, we have to change the model, rather than computing some fictive numbers which have no relation to reality. This, naturally, holds also for the Neyman-Pearson theory. As C. Daniel, a highly recognized applied statistician, once said: "We are told not to change the horses in the middle of the stream ...", but to continue along his line: If the old horses drowned already, we better use new ones. Cf. also [22].

2.4 More than two background assumptions

This subsection is an obvious generalization of 2.3. But if every B is not a whole belief system, but just a single parameter, Example 2 can be interpreted as a general inference method (related to minimum entropy methods).

Consider now a (finite or infinite) class of background beliefs or assumptions $B_1, B_2, ...$ with (prior) surprises, $b_1 = b(B_1), b_2 = b(B_2), ...$ In practice, we start looking only at the smallest b_i's; however, we have to be able to consider also larger b_i's, once an observation A is made, because now the (near) smallest $s(A|B_i) + b(B_i)$ will be of the greatest interest. And the B_i's with "very large" $s(A|B_i) + b(B_i)$ will be deemed "practically impossible".

Example 2: Let F be a measurable space with a set of parameters B_i and a collection of potential probabilities $P(A|B_i)$ for all measurable events A and the corresponding surprises $s(A|B_i) = -\log P(A|B_i)$. Let $b_i = b(B_i)$ the collection of apriori surprises if the B_i were declared to be true. The b_i may be a constant, or may be determined by a (subjective or objective) Bayesian apriori distribution, a likelihood, a belief function, or some other measure of the apriori "plausibility" of the B_i. Given an observation A, the aposteriori surprise of $B_i|A$ is $s_i := s(A|B_i) + b(B_i)$. All "small" values of s_i are entirely plausible, and we may just for convenience pick out the minimum or some similar quantity (perhaps depending also on "neighboring" B_i's, doing some local "smoothing"). But all "too large" s_i are ruled out as "practically impossible" (until perhaps – rarely – a very surprising future observation A_2 forces us to either scrutinize A and A_2 more closely or to revolutionize the order of the B_i, digging out hypothetical models not yet considered in practice so far).

For a qualitative and semi-quantitative description of such a set-up, with many practical examples, cf. [28] and [29].

3 Additional aspects

3.1 Cautious surprises and successful bets

This subsection is for the readers who either know the two concepts mentioned, or who may want to study the pertaining literature, and want to see its relation to the present paper. (Obviously, there is no space here to repeat the old theories.) The last three paragraphs contain sketches of related possible future research problems.

In [21] a function m between 0 and 1 (a kind of upper probability describing our lack of surprise about some event A) and two definitions are introduced, namely "cautious surprises" and "successful bets". Now we can put $s = -\log m$, and the property of cautious surprises is nothing but the minimization of the average surprise mentioned above.

When we linearize the logarithm of m, we obtain a linear theory with close relationships to other statistical concepts, especially bets, and the concept of successful bets has been worked out to some extent especially in [23, 24]. A very special case are the (in)famous fiducial probabilities [21, 26, 27].

Somewhat related may be the linearization of approximately linear theories, such as Choquet capacities in a local neighborhood (described, e.g., by the gross-error model or the total-variation model).

Another aspect may be the robustification of the potential surprises, by putting an upper bound on them. The need for this may be only moderate, since $m \log m$ is bounded on the unit interval; but the two factors may not be always so closely related.

The approximate or exact requirements of cautious surprises or successful bets may also help in the robustification of the Bayes theory, as in the "weighted Bayes' theorem" [25, Ch. 5.3], in which basically random weights are treated like fixed weights.

3.2 Ted and Fred: different background information

We now leave Ted considered in isolation and discuss informally some situations where more than one person is involved.

An important practical problem is that two (or more) persons – say, Ted and Fred – may have different background knowledge or beliefs, while some common ground, resulting in common or at least similar surprises, should be achieved, otherwise no general opinion, including no general scientific theory, would be possible.

A first step is to openly discuss the different background opinions of Ted and Fred, until (hopefully) some common agreement can be found.

But a frequent obstacle is that many opinions, or even many reasons for such opinions, are not conscious for either Ted or Fred. They may be subconscious prejudices, which perhaps only by some kind of hard detective work can be elucidated, for example, by auxiliary information given by Ted or Fred, or by their family, educational, sociological or religious background.

Even if the basic reasons for such disagreements can be brought out into the open, it may be that on certain points no agreement is possible. Then Ted and Fred still can "agree to disagree". An example is the technical staff for water, electricity etc. in West and East Berlin during the height of the Cold War, who had to cooperate in the divided city, and they did so productively, agreeing on the political disagreements, but making sure the city would still function.

4 Examples and further aspects

The last chapter provides some more examples of the rich variety of real-life situations which can be described within the framework given. I don't know all the literature, but I am not aware of a theory which, for example, does describe the zigzag in the example of 4.1 in an adequate way. Perhaps it is because most theories are only deductive, while in real life we need (also) inductive thinking. The formal framework may still have to be worked out further and refined, as the example in 4.2 shows. On the other hand, the example in 4.3 should be a relatively easy one also for other theories, as it involves no change of background; moreover, I found it decidedly useful; it may and should already exist somewhere, in some form or other.

The last paragraph offers many opportunities for further work. But at any rate, this paper, together with the two previous ones, provides a broad conceptual framework (if one wishes, even a quantitative one, as shown specifically in this paper) for describing how we can deal with incomplete knowledge and how we can learn in real life.

4.1 Updating of the background information

Updating information clearly is an important operation, which can change potential surprises considerably.

Let us assume Ted is going to visit Fred by train fairly late in the evening. Fred expects to meet Ted at the closest major station, perhaps a few minutes late, but hardly more than half an hour late; any much longer delay would be a big surprise. But then Ted calls that he is stuck somewhere, because of a serious accident on the route, and has no information on how long the delay will be. It is now conceivable that he cannot even reach the last local train. Later, he cites the experience of a fellow traveller that with this type of accident, the delay is usually around 2 hours. This would mean still reaching the last local train. Eventually, after the train moves again, two official delays become available, which both are somewhat below 2 hours, but differ by 20 minutes. The true arrival time is in between.

The consecutive updating of the background information changes the potential surprises, first to much less "knowledge", then to a more realistic expectation (although, as so often in life, not all discrepancies are cleared up).

4.2 Unexpected surprises

As mentioned above, some potential surprises are so unlikely to Ted that he does not even think of them. Sometimes he would consider them more plausible if his background information were updated by some additional information. Let us consider an example with various forms of surprises.

A married couple want to celebrate their wedding anniversary, with the husband secretly organizing it. First, they arrive at a high-level hotel, a fairly big surprise for the wife, but feasible. Then they get their room which turns out to be a ("the") historical room: almost everything like a hundred years ago: a big unexpected surprise for both of them. An excursion by horse-drawn carriage was only a moderate surprise for the wife, since such carriages exist in the area. But an excursion by public boat on the nearby lake was an "impossible" surprise for her, since she knew that such boats didn't exist; she needed the updating of her knowledge that in very recent years public boat connections had been introduced. But then the husband leads his wife, well-dressed and at a fixed time, not to the ordinary hotel elevator, but to the remote staff elevator; they go down and get lost in the subterranean floors; he finds the way again, and they walk amidst the rooms of the staff and end up in a little chapel where a priest performs a small private ceremony for their wedding anniversary. – It seems hard to formalize such surprises and the lack of any knowledge on the way there.

4.3 Informative short knowledge descriptions

Let us close with an example from field ornithology. There are many books on where to find which birds, but some of them I find rather unsatisfactory, either being not sufficiently informative, or not agreeing with my experience (or being even misleading). However, I discovered one book which, to my own surprise, I found very useful [2]: in its bird lists (each for a larger area), the abundance of every species was coded by just 3 symbols: `c` (for common), no symbol, or `r` (for rare). (To be more precise, there are also symbols for the season (summer, winter, migration, or year around) and sometimes for the altitude or other informative features.) Why are just these 3 symbols for abundance so satisfying, according to my experience?

Clearly, `r` means rare: not impossible, but each observation would be a big (pleasant) surprise, unless one knows and visits the restricted areas (if existing) where the species is not so rare. But in general it would be no surprise at all not to find the species, even after a long search. – And `c` means common: the species would be no surprise at all, and with a decently long search in the right habitat (and perhaps time of day, weather, etc.), it would be a big surprise not to find it. – No symbol means neither `c` nor `r`; it would be neither a surprise to find the species, nor a surprise not to find it. The species may be sparsely distributed, or regional, or temporal (e.g., during irregular invasions); a more detailed description of the "probability of encounter" ("Antreff-Wahrscheinlichkeit", cf. [19, 20, 22] would be too complicated on limited space. But two out of three categories are very informative. – I think we can use this set-up much more generally to distinguish the things we are pretty sure to happen, the ones we are pretty sure *not* to happen, and the ones we just don't know.

The method can be easily generalized to situations with more than two alternatives. We can describe profiles of potential surprises – and conditional surprises, given various backgrounds – in very complicated situations. Surprises imply assumed partial knowledge (that an event is not likely going to happen). Two very special cases are deterministic knowledge (all surprises infinite, except one being zero), and perfect knowledge of a probability space (the sum of the negative antilogarithms of all surprises of disjoint events being one), but obviously there are many more forms of incomplete knowledge.

Acknowledgments: I am grateful to W. Stahel for his help. – Besides the references given by the editors, two referees provided a thorough critical reading and a number of questions and suggestions which gave rise to a considerable enlargement of the original paper.

References

[1] C. F. Alchourron, P. Gardenfors, and D. Makinson. On the logic of theory change: partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50(2):510–530, 1985.

[2] P. Alden and J. Gooders. *Finding Birds Around the World.* Houghton Mifflin, Boston, 1981.

[3] A. Arnauld and P. Nicole. *La logique ou l'art de penser.* Paris, 1662. Reprinted 1965 by Fridrich Fromann Verlag, Stuttgart-Bad Cannstatt.

[4] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Philos. Trans. Roy. Soc. London*, A 53:370–418, 1763. Reprinted in *Biometrika* **45** (1958), 293–315, and in Pearson, E. S., Kendall, M. G. (eds.)(1970): *Studies in the History of Statistics and Probability,* 131–153, Griffin, London.

[5] J. Bernoulli. *Ars Conjectandi.* Thurnisiores, Basel, 1713. Reprinted in *Werke von Jakob Bernoulli, volume 1,* Birkhäuser Verlag, Basel, 1975.

[6] G. Boole. *An Investigation of the Laws of Thought, on which are founded the Mathematical Theories of Logic and Probabilities.* Macmillan, London, (1854) 1958. Reprinted by Dover, New York (1958).

[7] D. Brönnimann. Die Entwicklung des Wahrscheinlichkeitsbegriffs von 1654 bis 1718. Diplomarbeit, Seminar für Statistik, Swiss Federal Institute of Technology (ETH) Zurich, 2001. URL: `ftp://stat.ethz.ch/Masters-Theses/Daniel_Broennimann-Wahrschkeit.pdf`.

[8] M. E. G. V. Cattaneo. Likelihood-based statistical decisions. In F. G. Cozman, R. Nau, and T. Seidenfeld, editors, *Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications*, 2005.

[9] M. E. G. V. Cattaneo. *Statistical Decisions Based Directly on the Likelihood Function.* PhD thesis, ETH Zurich, 2007.

[10] A. Cournot. *Exposition de la théorie des chances et des probabilités.* Librairie philosophique J. Vrin, Paris, (1843) 1984.

[11] B. de Finetti. La prevision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré*, 7:1–68, 1937. English translation in Kyburg, H. E., and Smokler, H. E. (eds.) (1964) *Studies in Subjective Probability.* Wiley, New York. (2nd, enlarged ed. 1980).

[12] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, 38:325–339, 1967.

[13] A. P. Dempster. A generalization of Bayesian inference. *J. Roy. Statist. Soc.*, B 30:205–245, 1968.

[14] D. Dubois and H. Prade. *Theory of Possibility.* Plenum, London, UK., 1988. Original Edition in French (1985) Masson, Paris.

[15] R. A. Fisher. Inverse probability. *Proc. of the Cambridge Philosophical Society*, 26:528–535, 1930. Reprinted in Collected Papers of R. A. Fisher, ed. J. H. Bennett, Volume 2, 428–436, University of Adelaide 1972.

[16] I. J. Good. The probabilistic explication of information, evidence, surprise, causality, explanation, and utility. In V. P. Godambe and D. A. Sprott, editors, *Foundations of Statistical Inference*, pages 108–141. Holt, Rinehart, and Winston, Toronto, 1971. Reprinted partly in Good (1983).

[17] I. J. Good. *Good Thinking; The Foundations of Probability and Its Applications.* University of Minnesota Press, Minneapolis, 1983.

[18] I. Hacking. *The Emergence of Probability. A philosophical study of early ideas about probability, induction and statistical inference.* Cambridge University Press, Cambridge, 1975.

[19] F. Hampel. Artenliste vom Seeburger See 1955-1964 (unter knapper Berücksichtigung des Raumes um Göttingen). mimeographed manuscript, Göttingen, 23 pp., later reprinted, 1965.

[20] F. Hampel. Überwinterung und Verhaltensweisen der Beutelmeise (Remiz pendulinus) am Seeburger See. *J. für Ornithologie*, 107 (Vol. 3/4):359–360, 1966.

[21] F. Hampel. Some thoughts about the foundations of statistics. In S. Morgenthaler, E. Ronchetti, and W. A. Stahel, editors, *New Directions in Statistical Data Analysis and Robustness*, pages 125–137. Birkhäuser Verlag, Basel, 1993.

[22] F. Hampel. Is statistics too difficult? *Canad. J. Statist.*, 26(3):497–513, 1998.

[23] F. Hampel. On the foundations of statistics: A frequentist approach. In M. S. de Miranda and I. Pereira, editors, *Estatística: a diversidade na unidade*, pages 77–97. Edições Salamandra, Lda., Lisboa, Portugal, 1998. URL: `ftp://ftp.stat.math.ethz.ch/Research-Reports/85.pdf`.

[24] F. Hampel. An outline of a unifying statistical theory. In G. de Cooman, T. L. Fine, and T. Seidenfeld, editors, *Proc. of the 2nd Internat. Symp. on Imprecise Probabilities and their Applications, ISIPTA'01, Cornell University, 26-29 June 2001*, pages 205–212. Shaker Publishing Maastricht, 2000, 2001. URL: `ftp://ftp.stat.math.ethz.ch/Research-Reports/95.pdf`.

[25] F. Hampel. Some thoughts about classification. In K. Jajuga, A. Sokołowski, and H.-H. Bock, editors, *Classification, Clustering, and Data Analysis. Recent Advances and Applications*, Studies in Classification, Data Analysis, and Knowledge Organization, July 16–19, 2002, Cracow, Poland, pages 5–26. Invited keynote lecture, 8th Conference of the International Federation of Classification Societies, Springer, Berlin, 2002. URL: `ftp://ftp.stat.math.ethz.ch/Research-Reports/102.pdf`.

[26] F. Hampel. The proper fiducial argument. Extended abstract. *Electronic Notes in Discrete Mathematics*, 21:297–300, 2005.

[27] F. Hampel. The proper fiducial argument. *Information Transfer and Combinatorics*, LNCS 4123:512–526, 2006. URL: `ftp://ftp.stat.math.ethz.ch/Research-Reports/114.pdf`.

[28] F. Hampel. Upper and lower probabilities in real life. In *CD-ROM containing the Proc. 56th Session of the ISI, Contrib. Papers, Lisboa, Portugal*, 2007. URL: `ftp://ftp.stat.math.ethz.ch/Research-Reports/145.pdf`.

[29] F. Hampel. How can we get new knowledge? In T. Augustin, F. P. A. Coolen, S. Moral, and M. C. M. Troffaes, editors, *Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*, Durham, U.K., 2009. Durham University, Department of Mathematical Sciences. URL: `http://www.sipta.org/isipta09/proceedings/037.html`.

[30] F. Hampel. Nonadditive probabilities in statistics. *Journal of Statistical Theory and Practice*, 3(1):11–23, 2009. URL: `ftp://ftp.stat.math.ethz.ch/Research-Reports/146.pdf`.

[31] E. L. Lehmann. *Testing Statistical Hypotheses.* Wiley, New York, 1959. (2nd ed. 1986).

[32] A. Neumaier. Fuzzy modeling in terms of surprise. *Fuzzy Sets and Systems*, 135:21–38, 2003.

[33] B. Pascal. *Oeuvres complètes*, volume 34. Bibliothèques de la Pléiade, Paris, 1954. Edited by Jacques Chevalier, see also many more editions.

[34] E. S. Pearson and J. Wishart, editors. *"Student's" Collected Papers.* Cambridge University Press, 1958.

[35] K. Pearson. *The History of Statistics in the 17th & 18th Centuries; against the changing background of intellectual, scientific and religious thought.* Griffin, London, 1978. Lectures by K. Pearson at University College London 1921-1933, edited by E.S. Pearson.

[36] G. L. S. Shackle. *Decision – Order and Time in Human Affairs.* Cambridge University Press, Cambridge, (1961) 1969.

[37] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, N.J., 1976.

[38] "Student". The probable error of a mean. *Biometrika*, 6:1–25, 1908.

[39] P. Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, 1991.

[40] L. A. Zadeh. Fuzzy sets. *Inform. Control*, 8:338–353, 1965.