Ih – Hypotheses, Probabilities & Evidence

Dear: In the previous chapter, I encouraged you to abandon use of the words 'belief' and 'faith' and, instead, to say what you really mean. In this chapter, I want to show you that, in many if not most cases, the word 'belief' should be replaced by the word 'probability'. I then want to show you methods for estimating the probability of the truth of any hypothesis based on evidence; thereby, I hope to build your capabilities and confidence in holding your "beliefs" only as strongly as relevant evidence warrants.

That by 'belief' we sometimes mean 'probability' (a word derived from the Latin word *probus*, meaning "praised by a wise person") can be seen in the definition for 'belief' given in Webster's dictionary:

> **belief** **1.** the state of believing; conviction or acceptance that certain things are true or real **2.** faith, especially religious faith **3.** trust or confidence {I have *belief* in his ability} **4.** anything believed or accepted as true; especially, a creed, doctrine, or tenet 5. an opinion; expectation; judgment {my *belief* is that he'll come}

As in the case of essentially all dictionary definitions, the above definition for 'belief' reflects common usage. This common usage, however, reflects (at least) three unfortunate characteristics of most English-speaking people: 1) they're mired in supernaturalism, 2) they don't understand that "truth" can be ascertained only in religions and other games, and 3) they don't know how to transform (as given in the 5th definition) "an opinion; expectation; judgment" into estimates of associated probabilities. Thus, the above, 5th definition for 'belief' would be more complete in a form similar to:

> **belief… 5.** an opinion; expectation; judgment; probability {my belief is that he'll come; my belief is that there's a 75% chance that he'll come}

To see that, by stating a 'belief' we can be attempting to convey an estimate of a probability, try it out. Thus, Dear: Do you believe (better, "think") it'll rain today? "Well, the weather report said that the chance of rain was about 30%." [So: chance = probability = belief.] Do you believe (better "expect") the Dodgers will win tonight's game? "Well, I heard that in Las Vegas they're given 7:5 odds on the Dodgers." [So: expectation = odds = a measure of probability = belief.] Do you believe in God? "Well, I think I'll pass on that one, since it's not yet clear to me how to estimate probabilities in the absence of evidence." [Sorry, Dear, I couldn't resist the temptation of attributing those words to you!]

Let me try to “hammer the point home” with another example. Thus, suppose you asked your mother if she “believes” in God. I’ll bet (i.e., I estimate that the probability is close to unity or 100%) that your mother’s right-brain-stimulated response would be expressed with words such as “of course” or “certainly”. Now, Dear, in probability theory (as I expect you know) the words “of course” or “certainly” mean that the associated probability, p, is exactly unity: p = 1. But then, pestering your mother as you sometimes do (“Who, me?!”), you might then ask her: “But what do you estimate to be the probability that God exists?” Perhaps she’d answer [where by “perhaps” I mean: “I expect, with lower probability than ‘I bet’, that she’d answer”] something similar to: “What do you mean probability? I’m certain that God exists.” That would mean, that she takes for the probability of God’s existence to be unity, viz., p(G) = 1.

Relying on her technical training, however (i.e., relying on the stimulation of her left-brain’s analytical capabilities from her university training in chemistry), your mother might have learned that, in the real world, there’s no justification for assigning a probability of unity to any specific one of several possibilities. Thereby, she might answer differently, e.g., “I’m 99.99999999% certain that God exists”, i.e., p(G) = $1 – 10^{-10}$, if I counted all those 9’s correctly! Correspondingly, since the sum of probabilities for all possibilities must sum to unity [from the meaning of ‘probability’, because, with certainty (i.e., a probability of unity), at least one of the possible outcomes must occur], then your mother would simultaneously be allotting a probability of 1 chance in 10^{10} to the possibility that God doesn’t exist. Thus, she’d be assuming p(¬G) = 10^{-10}, where I’ve again used the symbol ‘¬’ for the word ‘not’, and where I’ve satisfied the requirement that either God exists or doesn’t exist, i.e., p(G) + p(¬G) = 1.

In any event, Dear, I hope you see that, any statement of any “belief” (about the weather, about the outcome of some undertaking, about “the nature of anything”, about the existence of “whatever”) can be a statement about one’s estimate of relevant probabilities. Which then leads to the obvious question that’s simple to ask (but in many cases not simple to answer!): What should we do to try to make our estimates of such probabilities (i.e., such “beliefs”) rational or reasonable or reliable? That is, Dear, obviously a distinction should be made between rational and irrational beliefs (or rational and irrational estimates of probabilities). And as you might expect, the distinction is derived from one’s capabilities to estimate probabilities.

In earlier chapters, I tried to show you the importance of being able to make realistic (or reasonable or reliable) estimates of probabilities. For example, I tried to show you in Chapter **H2** that to hope is to gamble [in which one seeks to optimize some investment's expected value = (value – investment) x *probability*]. Also, I tried to show you in Chapter **D** that to make rational decisions, you need reliable estimates of relevant probabilities (because people normally choose the option with minimum risk and maximum reward, i.e., the option with greatest "utility" or "expected value" = value x *probability*). That is, Dear, for the hundreds of decisions that you'll need to make every day, your estimates of relevant probabilities will be important – sometimes, critically important. Therefore, it's important that you develop capabilities to make rational, realistic, and reliable estimate of probabilities.

In this chapter, however, I want to emphasize that, with competence in estimating probabilities, you'll then be able to adopt "beliefs" that are more realistic. For example, if as a scientist you had developed some hypothesis, then when new data are obtained (that either support or undermine your hypothesis), you would be able to estimate how you should rationally change the confidence (or "belief") in your hypothesis. Alternatively, if you (in your gullibility as a child) had accepted your parents' ideas about God and if, as you grew older, you found evidence that either supported or distracted from their "hypothesis" about the existence of God, then as you gained more "discriminating power" (i.e., as you developed more competence in estimating probabilities), you'd be able to evaluate appropriate changes to your "belief", i.e., your estimate of the probability that your parents' proposition is true.

Consequently, just as in earlier chapters when I started to describe decisions and cautioned you to try to adopt realistic hopes, I'm again back to the question of how to estimate probabilities realistically. In all cases, what should be done to estimate probabilities is to use "common sense." In some cases, however, common sense actually isn't so common: witness the many people whose estimates for the existence of any god leads them to "believe" that such "animals" actually exist!

In an earlier chapter in this group of **I**-chapters (namely, **Ic**, entitled "Constraining Ideas"), I tried to encourage you to constrain your imagination, when appropriate. I ended that chapter with the quotation from Simon Ewins:

> God is a perfect example of the kind of aberration that can result from an untrained intellect combining with an unrestrained imagination.

In this chapter, I want to try to help you train your intellect, so you'll increase your competence in basing your rational decisions on evidence – plus application of a little common sense.

Of course, I'd be pleased if you used not just "a little" but "a lot" of common sense, but to do the latter, it would help if you first completed your Ph.D. thesis on a topic that would demonstrate your competence in Bayesian probability theory! But actually, as I'll show you at the end of this chapter, such an approach wouldn't guarantee success. In particular, I'll show you a case in which a recent Ph.D. theoretical physicist published a silly book in which he uses Bayes' theorem to convince himself that the probability that God exists is 2/3. Unfortunately, however, although he may be competent as a theoretician, he seems incompetent as an experimentalist, since he demonstrates that he doesn't have a clue about what "evidence" means!

But I'm getting ahead of myself. To try to constrain one's imagination, to try to be more logical, to try to estimate probabilities more "reasonably", one might hope to start with Aristotelian logic. Unfortunately, however, there's substantially more to reasoning (and to "common sense") than Aristotle saw. In an earlier chapter (**Ib2**, entitled "Basic Ideas in Logic"), I suggested that Aristotelian logic has some severe limitations – to which I would return (your patience assumed). If your patience persists, then in subsequent chapters (e.g., in **R**, dealing with Reasoning, in **T** dealing with Truth, and in **U** dealing with Uncertainties), I'll show you some details about limitations of Aristotle's ideas. In this chapter, however, I want to illustrate "just" some general features of limitations on his analysis of logic, a general description of which is that his "true or false" option is far too restrictive.

To see what I mean, Dear, please consider the following figure, which appears as Figure 3.1 in an excellent 1999 report by Giulio d'Agostini of the University of Rome and entitled *Bayesian Reasoning in Physics: Principles and Applications*.[1]

---

[1] Available at http://public.lanl.gov/kmh/course/bayesian.html. This report was published as CERN Yellow Report 99-03, July 1999 (vi + 175 pages) under the title *Bayesian reasoning in high energy physics. Principles and applications.* The reference that d'Agostini gives as the original source of the figure is "[28]: B. de Finetti, *Probabilita*, entry for *Enciclopedia Einaudi*, 1980."

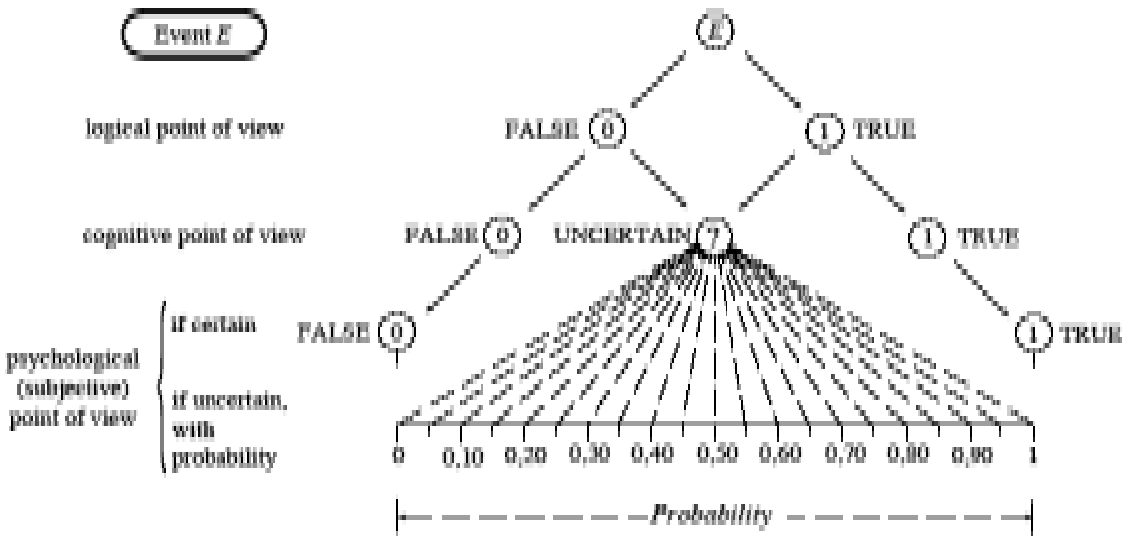


Figure 3.1: Certain and uncertain events [28].

If you will examine this figure starting at the top left, an "Event E" is assumed to exist, whether it's what we usually think of as an "event" (in space and time) or it's some statement or claim (such as "the sky is blue" or "God exists"). At the top right, this event, E, sits at the top of a "pyramid of possibilities".

On the left, next lower is the prescription "logical point of view" (in particular, Aristotelian logic). On the right at the same level, the first level down from the top of the pyramid of possibilities, are only two options: either the event is FALSE (numerical value of zero) or TRUE (numerical value of unity). Those are the only two possibilities in "classical" or Aristotelian logic (and the basis of "Boolean algebra", with its 0 or 1 options used in "computer logic").

At the next lower level, from the "cognitive point of view", the picture is starting to "get real". That is, to the right, on the "pyramid of possibilities", is shown the case that's much more common: not that some event (or claim) is just TRUE or FALSE, but that, between the two, is a wide range of possibilities that, without further information, are UNCERTAIN.

Finally, at the lowest level on the left is the "psychological (subjective) point of view", i.e., what each of us must deal with when considering any event.

For this "psychological [or personal] point of view", on the left are shown two possibilities, depending on "if [we're] certain" (that the event is either TRUE or FALSE) – which, in reality, essentially never occurs – or "if [we're] uncertain". Then, for all those cases of uncertainty, the best we can do is associate probabilities with the event, which (as shown by the numbers at the bottom of the "pyramid of possibilities") range from 0 (certainly false) to 1 (certainly true).

Thereby, Dear, consider how far it's necessary to go beyond Aristotle's ideas, if we are to develop better ideas of both reality and how we actually think about reality. Fortunately, however, each of us has learned (more or less well) how to "handle" uncertainties; it's commonly called "common sense". We manifest this "common sense" either with our right-brain's synthesis capability or our left-brain's analysis capabilities (or both).

Actually, Dear, to include cases in which Right Brain dominates in defining "common sense", a "ground layer" to the above "pyramid of possibilities" should be added. That is, if we don't assign our left-brain to undertake the tasks of evaluating probabilities quantitatively (i.e., determining their numerical values between 0 and 1), then Right Brain can summarize our feelings about any event or claim with a range of emotions that could be shown at the base of the pyramid. This range of emotions can extend from laughter (maybe accompanied by a statement such as "Come off it!" or "Get real!") or jeers (maybe with a statement such as "Okay, let's see YOU jump over the moon!"), through snickers (recall the "snicker test"), amusement ("Ha – do tell!" or maybe "You've gotta be kidding!"), puzzlement ("Wow – that's weird!"), interest ("Hmmm, I'll need to think about that"), and so on, all the way out to affirmation (e.g., "Yah – right on!").

But in this chapter, rather than dwell on emotional assessments of "events", what I want to do is dig deeper into left-brain's analysis capabilities for estimating probabilities quantitatively. Supporting that goal is the foundational statement (which I've quoted before) by one of the "fathers" of probability theory, Pierre-Simon Laplace (1749–1827): "The theory of probabilities is at bottom nothing but common sense reduced to calculations." In this chapter, therefore, I want to show you how progress can be made making real-world decisions about the ideas or hypotheses that you may want to adopt (and others that you may want to reject) by using evidence and some common-sense estimates of probabilities.

In general, however, the task of reducing common sense to calculations isn't easy – in the main, because our "common sense" is actually amazingly complicated. To begin to see what I mean, consider the following statement by E.T. Jaynes (1922–1998) dealing with "plausible reasoning":[2]

> The syllogism is the standard example of deductive reasoning: if A is true, then B is true; A is true; therefore, B is true (or it's inverse: if A is true, then B is true; B is false; therefore, A is false). This is the kind of reasoning we'd like to use all the time, but unfortunately, in almost all situations… we don't have the right kind of information to allow this kind of reasoning. We fall back on weaker forms: if A is true, then B is true; B is true; therefore, A becomes more plausible. The evidence doesn't prove that A is true, but verification of one of its consequences does give us more confidence in A.
>
> Another weak syllogism, still using the same major premiss, is: if A is true, then B is true; A is false; therefore, B becomes less plausible. In this case, the evidence doesn't prove that B is false, but one of the possible reasons for its being true has been eliminated, and so we feel less confident about B. The reasoning of a scientist, by which he accepts or rejects his theories, consists almost entirely of syllogisms of the second and third kind…
>
> A still weaker form [of syllogism is]: if A is true, then B becomes more plausible; B is true; therefore, A becomes more plausible… This [an example is given] shows that the brain, in doing plausible reasoning, not only decides whether something becomes more… or less plausible, but it evaluates the degree of plausibility in some way. And it does it in some way that makes use of our past experiences as well as the specific data of the problem we're reasoning on… This reasoning process goes on unconsciously, almost instantaneously, and we conceal how complicated it really is by calling it common sense.

Which then leads me to the need to make a decision. One option is to show you, next, details of how "common sense" can be "reduced to calculations". By far the best way to do that (following Jaynes' book) is develop the theory of probability from a few "basic axioms", which then lead to Bayes' theorem, and then show how the theory can be applied in a number of practical problems, so you'll "get a feel for it" and thereby conclude something similar to: "This is trivial stuff; it's just common sense!"

---

[2] As you can find using a "search engine", Jaynes' excellent online book *PROBABILITY THEORY with Applications in Science and Engineering* can be found at several locations on the web (e.g., at http://omega.albany.edu:8008/JaynesBook.html). The web version is "a series of informal lectures" and is described as "a fragmentary edition of June 1974." Subsequent to Jaynes' death, G.L. Bretthorst successfully undertook the challenging task of transforming Jaynes' "informal lectures" into the book *Probability Theory: The Logic of Science* by E.T. Jaynes, G. Larry Bretthorst (Editor), which was published by Cambridge University Press in 2003 and is available at, e.g., amazon.com. An "unofficial errata and commentary" for this book is at http://leuther-analytics.com/bayes/jaynes/.

In an earlier version of this chapter, I tried that approach, but I became quite discouraged with the length to which the chapter grew, even trying to show you why the "basic axioms" of probability conform to common sense! Consequently, I've chosen another option, namely, to show you just how some "common sense" can be used directly to estimate probabilities (at least in some simple cases) – and then, I'll show you at least an outline of how Bayes' method can be used. In a later chapter (in **T2**, dealing with "Truth" and Understanding) I'll show you additional details about Bayes' method and how it's used in applying the scientific method.

Now, in earlier chapters (e.g., in **H2** dealing with "Hope"), I sketched a little about how to estimate probabilities for some simple cases, such as when some expert advice is available (e.g., the probability that it'll rain today or the probability that your suede jacket will be damaged if you get it wet), when some symmetry is apparent (e.g., in tossing a coin or throwing a die and similar), or when the outcome of a large number of trials can be examined (either in your head or in reality). In this chapter, therefore, let me show you at least a little about how to estimate probabilities in more complicated cases, when symmetry arguments, expert advice, or repeated trials are unavailable, impossible, or even undesirable – which is probably the case for most of our decisions. I'll start with some simple, hypothetical examples, to try to make explicit what you normally do when you apply common sense.

1. Estimating the probability that the Sun will rise tomorrow.

For my first example (following an example given by Richard Price, who after Bayes' death in 1761 communicated his paper to the Royal Society of London), suppose you were "brand new" to this world and you wanted to determine the probability that the Sun would rise the next day (for, as far as you knew, a 'day' was defined only by any 24 hour time-period). During the first night, with your not knowing if the Sun would rise the next "morning", suppose you decided (by "pulling a number out of your hat") that there was a 50-50 chance that the Sun would rise.[3]

---

[3] By the way, Dear, thereby you'd be acknowledging that you had no information about the system – which is tantamount to your assuming that the entropy of the system (i.e., your ignorance of its "internal workings") was at a maximum. And to be able to add that (important!) statement, thanks are due Claude Shannon, 1916–2001, the "father of information theory" (and of the field of Artificial Intelligence), who should have been awarded a Nobel prize in physics, but apparently he was considered to be a mathematician (and there's no Nobel prize in math) or he didn't have a sufficient number of "political connections" to be nominated for the Nobel prize in physics (which he undoubtedly deserved).

Further, suppose you decided that you'd keep track of the results of your test of the hypothesis that the Sun would (or would not) come up each day (i.e., keep track of the data or evidence) by putting a white ball in a bag every day that the Sun came up and a black ball in the bag every day the Sun didn't. On that first night, you put one white and one black ball in the bag, to indicate your initial guess at the probabilities (i.e., that the probability that the Sun would rise = $p_o(S) = ½$ [= (1 white ball) ÷ (a total of 2 balls)] and for the initial probability that it wouldn't rise = $p_o(\neg S) = ½$, in which ¬S means "not Sun" and the subscript "o" indicates that this is your "initial" or "zeroth" estimate (also called "the prior probability" or the "*a priori* probability" = "before the fact" probability).

Come the first morning, when the Sun came up, you then methodically placed a white ball in the bag (bringing the total contents of the bag to a total of 3 balls, 2 white and one black) from which you could then calculate your revised probabilities (with subscript "1" to identify your first "experimental result"): $p_1(S)$ = "favorable outcomes" / "total trials" = 2/3 and (of course) $p_1(\neg S) = 1/3 = 1 - p(S)$. Thus, the first experimental test of the hypothesis that the Sun rises each morning yields $p_1(S) = 2/3$. Thereby, you revised your estimate of the probability that the Sun would rise any morning – you revised your hypothesis – based on the evidence.

Incidentally, if you prefer, you could calculate the new "odds" for the Sun rising as $O_1(S)$ = "favorable outcomes" to "unfavorable outcomes" = 2:1 = 2/1 = 2. From which, by the way, and from continuing to use this "frequency interpretation of probabilities" (because a good general rule is: use what works!), it's easy to see how odds, *O*, can always be calculated from probabilities (or *v.v.*). Thus, start from the definition

Odds (of "winning") = {"favorable outcomes" / "unfavorable outcomes"} ;

next, divide numerator and denominator by "total outcomes" = N; then, upon recognizing the "frequency definition" of probability is "favorable outcomes"/ ["total outcomes" = N], and again using p(S) = probability that the Sun will rise and p(¬S) = probability that the Sun won't rise, the result is

$$O(S) = p(S) / p(\neg S) = p(S) / [\ 1 - p(S)\ ] .$$

This (general result) shows that, for probabilities small compared to unity (i.e., $p << 1$), then odds are essentially the same as probabilities, whereas for probabilities near unity (i.e., $p \rightarrow 1$), the odds become very large; thus,

$$O(\mathrm{S}; p << 1) \sim p(\mathrm{S}) \;\text{ and }\; O(\mathrm{S}; p \rightarrow 1) \sim 1/[1 - p(\mathrm{S})] = 1/[p(\neg\mathrm{S})] ,$$

in which I've used the symbol "~" to mean "asymptotically approaches".

If you persevered in your experimental test of the hypothesis for a full year, then you'd have become quite confident in the validity of the hypothesis that the Sun rises each morning, concluding (from the number of white and black balls in your bag) that $p_{365}(\mathrm{S}) = 365/366 = 0.997$ and $p_{365}(\neg\mathrm{S}) = 1/366 = 0.003$ (to three decimal places), or $O_{365}(\mathrm{S}) = 365/1 = 365$ and $O_{365}(\neg\mathrm{S}) = 1/365 = 0.003$ (to three decimal places). Upon seeing your results, someone unfamiliar with estimating probabilities might say "I *believe* that the Sun rises every morning", but you could correct him by saying: "Well, it's better to say that the evidence-to-date suggests that the probability that the Sun rises every morning is 99.7%."

If that "believer" persisted, if he went so far as to say "I'll give you 1000:1 odds that the Sun will rise tomorrow," then you might want to take the bet! Thus, based on your probability estimate (from a year's worth of data), then for every \$1 you bet on the chance that the Sun would fail to rise the next morning, your "expected return" (= value x probability) would be \$1000 x 1/365 ≃ \$3. That would seem to be quite a "rational bet" – unless, of course, you had the sneaking suspicion that the first ball in the bag was a mistake, concluding that its presence in the bag was skewing your estimate for the probability. Thus, if you hadn't started with one black (and one white) ball in the bag, then after collecting a year's worth of data, your estimates would be $p_{365}(\mathrm{S}) = 365/365 = 1$ (i.e., certainty) and $p_{365}(\neg\mathrm{S}) = 0/365 = 0$, and generally speaking, betting against certainty is irrational!

The moral of which, Dear, is: don't bet unless the odds offered are not only better than the odds based on your estimated probability but also better than the odds based on probabilities within some "confidence limit" of those probabilities, i.e., always be worried that your prejudices (i.e., your "pre-judgments" or your estimates of "prior probabilities") distort your estimate. To perform such an estimate, however, you'll need to do more than just count black and white balls in your bag – and the best way is to use Bayes' theorem (as I'll outline later in this chapter).

And maybe it would be useful if I dug up another moral from the above. Thus, Dear, suppose that, rather than meeting a gambler (who offered you 1000:1 odds that the Sun wouldn't rise the next day), you met a "Sun god worshiper". Suppose he explained to you "Of course the Sun rises each morning – because every night, we hold a prayer meeting at the temple, confess our sins, and make offering to the great Sun God, Sol." "Offerings," you responded somewhat warily (since you had heard some quite strange noises, nightly coming from the temple), "what sort of offering?"

"Well," he answered slowly, "our first great prophet, Muses, communicated with Sol and learned that the desired offering would be the life of a beautiful young virgin girl; so, each night our ancestors impaled one on the altar." "Oh," was your first response, adding "didn't that policy make it rather difficult for the men of your tribe to find marriageable women?" "Indeed it did," he responded in all seriousness, "but more virgins were found in neighboring tribes – and furthermore, our second wise prophet, Muh Hamad, communicated with Sol and learned that a better sacrifice would be the warriors of other tribes." "Ah ha," you sarcastically responded, "I can certainly see the wisdom in that alternative! " "True," said he, "but when enemies became difficult to capture and we became desperate that Sol would be offended, refusing to rise in the morning, our third wise profit, Smuth, communicated with Sol and learned that a sufficient offering would be for each of us to bring with us to the temple a tenth of what we produce, and to this day, Sol has been satisfied with our tithes, never failing to rise." And although several morals of that story probably come to you, Dear, let me express one in a manner that you might not have considered: when you examine evidence from any experiment, try (hard!) to understand what the experiment really reveals.

2. Estimating probabilities in a legal trial.

But enough of that. Now, consider a more practical example (than trying to estimate the probability that the sun comes up each morning – which actually was a case differing little from the usual "frequency method" for estimating probabilities). For this example, suppose you're called to serve on a jury, charged with passing judgment on an accused murderer. You are dutifully "sworn in" and you agree with the basic premiss of our judicial system that the accused is assumed to be innocent "unless proven otherwise, beyond a reasonable doubt."

Actually, though, viewed from the theory of probability, that "basic premiss of our judicial system" isn't stated well. When estimating probabilities, we essentially never start by setting a value for any probability to be exactly zero (i.e., that some claim is definitely false) – and for that matter, we rarely start with a value of exactly unity (i.e., that some claim is definitely true), because if either were so, that would be the end of the matter! That is, if the probability that the accused is guilty were exactly zero, then he should immediately be exonerated, because he's definitely innocent.

In reality, however, the presumption of innocence means, not that the person is "truly innocent", but that he's as innocent as anyone else. Thus, if there were 10^{10} people in the world, then the initial assumptions are 1) that someone did the deed, and 2) that the person accused is assumed to be just as innocent as anyone else, i.e., your initial estimate for the probability that he's guilty, say $p_o(G)$, is to be taken as the same as for everyone else: $p_o(G) = 1/10^{10} = 10^{-10} = 0.0000000001$.[4]

Now, let the trial begin. As a jurist, your job is to revise your estimate of the probability that the accused is innocent, $p(I) = 1 - p(G)$, provided (of course) that the evidence warrants such a revision. And notice, Dear, both that this is going to be a "one shot affair" [i.e., there won't be any "repeated trials" {literally – at least if there's a law on the books against "double- (or multiple-) jeopardy"}] and that you'll need to do the calculations in your head (using common sense), because there's no bag to hold appropriate black and white balls – and usually no computer to apply Bayes' method!

Well, Dear, similar to a lawyer, I could drag this trial on for a long time, but in the interest of space, let me quickly suggest what you would do as a member of the jury, applying common sense. Upon learning that the accused was seen near the crime scene, you might think: "So what? In this city, there could have been 10,000 people in that neighborhood that night; so, such evidence only boosts the probability that the accused is guilty up from 1 in 10 billion to 1 in 10,000, i.e., $p(G) \simeq 10^{-4}$."[5]

---

[4] Whoops. Here I'm using "G" to mean "guilty", whereas earlier in the chapter, I used "G" to mean "God" (and later in this chapter, I'll do it again). Sorry about that, Dear, but there's a limited number of letters in the alphabet!

[5] Dear: In that estimate, I used the sign "≃" (to mean "approximately equal to") in part because I don't want to show you the details (using Bayes' theorem) of how to calculate the revised probability more accurately, but especially because "common sense" says that the "evidence" (and the premisses) don't warrant such accuracy. That is, Dear, in most cases, common sense will tell you that, at least to start,

Next, suppose you heard evidence that the accused hated the victim and had even threatened his life. But then, the defense attorney showed that the victim was a real SOB, hated by hundreds of people; so, you weren't overly impressed that the accused seemed to have a motive. Apparently many people did – and just because a person has a motive, doesn't mean that the person is guilty of a crime. But it was evidence of a sort; so, you revised your estimate of the odds of his guilt to be, say, 1 in 100, i.e., $p(G) \simeq 10^{-2}$, but you did so without much conviction (i.e., with little confidence in the reliability of that estimate).

Then, some more compelling evidence was presented: you heard a seemingly reliable witness say that, she heard what sounded like two gunshots and then saw the accused running out of victim's room, with blood on his face. Given that evidence, suppose you thought: there's only one chance in 100 that he's innocent. Incidentally, notice that, thereby, you'd (probably) switch from making estimates (in your head) not of the probability of his guilt [which started at $p(G) \simeq 10^{-10}$] but of the probability of his innocence, now estimated to be $p(I) \simeq 1/100 = 10^{-2}$, of course with $p(G) = 1 - p(I) = 0.99$.

But the prosecutor hasn't finished her case: she next shows that the gun that killed the victim (which was found in the room) belongs to the defendant and has his fingerprints on it. With that information, you revised your "guesstimate" of his innocence to $p(I) \simeq 10^{-4}$ or $p(G) \simeq .9999$. Finally, she shows the police photos of scratches on the defendant's face and skin tissue under the victim's fingernails, with DNA matching the defendant's. You then make your final estimate of his innocence (with subscript "f" for "final") to be $p_f(I) \simeq 10^{-10}$ or $p_f(G) \simeq .9999999999$.

Well, Dear, although I dragged that out longer than I wanted (or expected!), maybe it would be useful to call your attention to a few features of the result. For one, notice that during the course of the trial, your estimate of the probability of the defendant's innocence (or guilt), i.e., your "belief" in his innocence (or guilt) changed enormously: from an initial $p_o(G) = 1 - p_o(I) = 10^{-10}$ to a final $p_f(I) = 1 - p_f(G) = 10^{-10}$. That's an astoundingly huge variation!

---

calculate just "ball-bark figures" – by which people mean that, to play the game, you should first make sure you're in the right ball park! Also, I haven't added subscripts on "p(G)" to indicate it's your zeroth, first, second, etc., estimates of the accused's guilt, because the notation "gets to be a pain"!

Written in terms of the "odds" of his guilt, for which (as I showed you near the end of the example dealing with the probability that the Sun would rise the next day) $O(\mathrm{G}) = \mathrm{p(G)}/[1 - \mathrm{p(G)}] = \mathrm{p(G)}/\mathrm{p}(\neg\mathrm{G})$, and using the result that for small p(G), $O(\mathrm{G}) \rightarrow \mathrm{p(G)}$ and for p(G) near unity, $O(\mathrm{G}) \rightarrow 1/\mathrm{p}(\neg\mathrm{G}) = 1/\mathrm{p(I)}$, then the range in your estimate for the odds of his guilt varied from $O_o(\mathrm{G}) = 10^{-10}$ to $O_f(\mathrm{G}) = 10^{10}$, i.e., 20 orders of magnitude! In such cases, there's no need to perform elaborate calculations of probabilities (e.g., using Bayes' method), and in fact, there's little justification, since all estimates are approximate. In such cases, "ballpark figures", obtained by doing "common-sense calculations" in your head, are quite adequate.

You might also want to think about the idea that the accused was found guilty "beyond a reasonable doubt". That's certainly a "subjective" criterion – showing the need for application of more "common sense". Besides, it's also quite a misleading statement (as misleading as the statement that the accused is initially assumed to be innocent, rather than "as innocent as anyone else"). And I say that "beyond a reasonable doubt" is misleading, because it fails to define "reasonable". Some "doubt" should always exist, but "reasonable" is to be judged with respect to what objective?

I suspect that a common-sense answer to that question would incorporate two ideas: 1) that "society as a whole" benefits if a lot of law breakers aren't "running around loose" (free because of some doubt about their guilt), and 2) such benefits for the rest of us must be balanced by the undesirable consequences (for each of us) if innocent people (such as we, of course!) are thrown in jail (or worse) because of unsubstantiated allegations. Where one "draws the line" between those two extremes, what one decides to be "reasonable" with respect to the objective of putting most criminals in jails while not incarcerating too many innocent people, of course depends on the common sense of each juror to decide, but for me (and apparently for the author of the probability book that stimulated this example, i.e., E.T. Jaynes), I'd draw a line somewhere around $\mathrm{p(I)} = 10^{-4}$ [or p(G) = 0.9999], by which I mean: I'd prefer if one in ten thousand guilty people were set free rather than if a larger fraction of innocent people were convicted.[6]

[6] The famous Jewish philosopher and medical doctor Moses Maimonides (1135–1204) argued: "It is better and more satisfactory to acquit a thousand guilty persons than to put a single innocent one to death."

Yet, in the case illustrated above, with a final estimate of $p_f(I) \simeq 10^{-10}$, the result was, in fact, off by 20 orders of magnitude! What actually happened on that fateful night was that, although the accused went with his gun and the intent to murder (and therefore was guilty of attempted murder), there was a struggle (during which the victim gouged the face of the accused), and the gun went off. All of which scared the accused so much that he dropped the gun and ran. Then, upon hearing a single shot, the woman (the victim's jilted lover – and the witness who said she heard two gun shots!) went into the victim's apartment, found him dazed, picked up the gun (with her gloved hand), and shot her ex-lover.

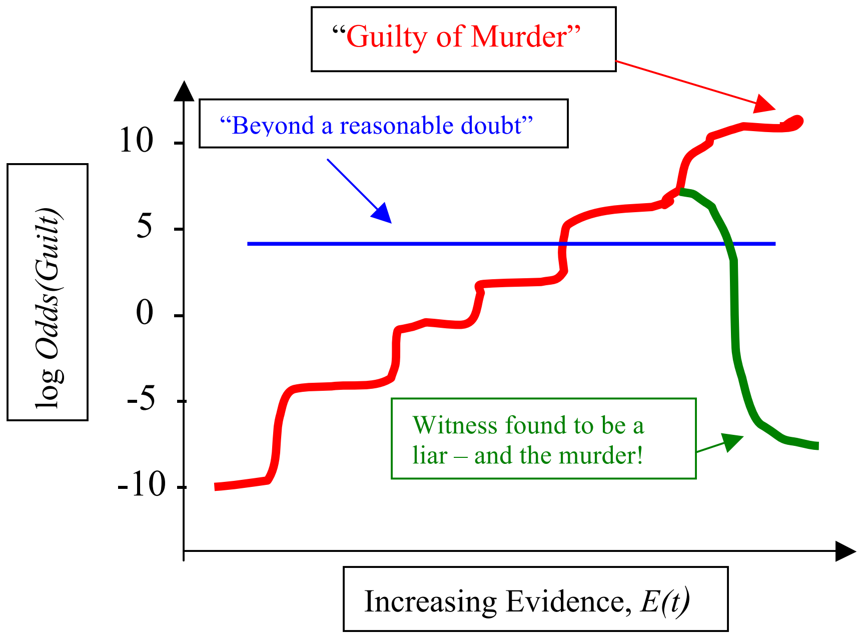
If you wanted to graph such change in belief as evidence accumulated – even if the plot were only "schematic" – then a linear graph wouldn't be very revealing. Thus, if you plotted the probability [say of his guilt, p(G)] on a linear scale, running between zero and unity, then with increasing evidence, the plot of p(G) would appear to be what's called a "step function": barely different from zero [when p(G) is in the range from 10^{-10} to about 10^{-2}] and then "stepping up" (in a jump!) to be barely different from unity [when p(G) is in the range of from about 0.99 to 0.9999999999]. Rather than a linear scale for p(G), consequently, it would be better to use a logarithmic scale for p(G), i.e., if a numerical value of p(G) is written as a base number (say *b*) raised to some power (say *c*), i.e., if $p(G) = b^c$, then plot only the "exponent" *c*, known as the logarithm of p(G) to the base *b* (and abbreviated with the word "log"): $\log_b p(G) = c$.

Given that all the probability estimates given above were written as powers of the base 10 [starting from "presumed innocence", with $p_o(G) = 10^{-10}$], then it would obviously be convenient to use base-10 logarithms, which are customarily written as just "log" (with "base-10" understood). Further, since as p(G) approaches unity, its log doesn't show much change [as p(G) changes only from, say, 0.99 to 0.9999999999, while p(I) correspondingly varies from 10^{-2} to 10^{-10}], it's therefore more revealing to plot, not the log of p(G) [or of p(I)], but the log of the odds, $O(G) = p(G)/[p(I)]$, i.e., plot

$$\log O(G) = \log p(G) - \log p(I)\ .$$ [7]

Illustrative schematics are given below, with odds plotted as a function of "increasing evidence with time", *E(t)*.

[7] Here, Dear, I've used log [x/y] = log x – log y, which you can check by looking again at exponents.



All of which then supports the two "cautions" that I raised after going through the first example (dealing with the probability that the Sun would rise the next day): 1) Be careful to understand what the evidence really reveals, and 2) Try to assess what confidence should be placed in probability estimates, i.e., what's the probability that your probability estimates are reliable?! To those two cautions I'd now add a third: 3) Be careful of the existence of hidden hypotheses. In this case, for example, you assumed that the witness was reliable, whereas in reality, she was the murderer.

3. Estimating probabilities using Bayes' method.

In my third of these examples of how to estimate probabilities, I'll first show you how common-sense calculations can be done in you head and then (finally!) show you how Bayes' method can be used to estimate probabilities more accurately. The basis of this example is one given in the CERN report by Giulio d'Agostini's that I referenced in the first footnote. In turn, his example is similar to the example that initially stimulated Bayes to develop his method: prior to Bayes, probabilities were used to predict probable outcomes (especially in games of chance); Bayes tackled the "inverse problem": he wanted to know what could be inferred about details of a particular game of chance from knowledge about the outcome.

For this next example, suppose you met with a friend for lunch one day each week. One week she suggested that who pays for the lunch should be decided by tossing a coin. Always game for a game (!), you took her up on it, called "heads", and lost. Okay, you win some and you lose some. If the coin were "fair", there'd be an even chance that you'd win or lose: the odds are 50:50 (or 1:1!), i.e., $p_o(\text{Win}) = p_o(W) = ½$ and $p_o(\text{Lose}) = p_o(L) = 1 - p_o(W) = ½$. Next week, the same thing happened – you called the toss again, and again you lost. Tough break, kid, but then, with the chance that you'd lose the first time = ½ , the chance that you'd lose two times in a row is $p_2(L) = ½ \times ½ = ¼$. Surely next time you'll win! Sorry, kid, although the probability that you'd lose three times in a row is $p_3(L) = (1/2)^3 = 1/8$, you lost again. But you're getting suspicious that "somethin' ain't right": with $p_3(L) = 1/8$, there was only 12.5% chance that you'd lose – but you did. Yet, that's nothing: convinced that soon your luck would change, you continued on for seven more weeks, losing every week, and although by that time the probability that you'd lose (10 times in a row) was $p_{10}(L) = (1/2)^{10} = 1/1024$ (i.e., $p_{10}(L) < 0.001$), yet, still you did. Common sense (and those common-sense calculations) then convince you that "somethin' ain't right".

So, you begin thinking about the wisdom of confronting her with your suspicion that (somehow) she's cheating (and therefore, that she isn't worth having as a friend). Before confronting her, however, you decide to refine your calculations using Bayes' method – to see how reliable are the "guesstimates" that you did in your head, using "just" common sense. Bayes' method deals with probabilities of compound statements (or sequential events), such as the probability that it will rain AND that you'll get your jacket wet, or the probability of the sequential event that your "friend" wins the first coin toss AND then the second – or, for that matter, the chance that she wins n-tosses in row.

Conditional probabilities (unsurprisingly) are those that depend on other conditions. They're invariably written with a "vertical bar" (i.e., " | "), which is read "given". Thus, p(A|B) is read "the probability of A, given B". In fact, all probabilities depend on some conditions, even though the conditions might not be stated. For example, when a weather forecaster states that the probability that it'll rain tomorrow is 30%, i.e., p(Rain) = 30%, he omits mentioning that his forecast depends on a number of conditions (I'll label them with C's), such as the condition that (C_1) the Earth will be here tomorrow, that (C_2) there'll be a tomorrow, and whatever other conditions (C_i) are implied. Thus, the probability that it'll rain tomorrow should be written as p(Rain|C_1, C_2… C_i…), but it's assumed that everyone knows that the probability statement is "conditional"; so, the weather forecast is given as, for example, just p(Rain) = 30%.

The fundamental rule for calculating compound (or sequential) probabilities [which you can take as a definition, which you can prove starting from Kolmogoroff's three axioms of probability theory (as you can find in essentially any elementary text on probability theory), or which you can just reason out using common sense] is that the joint probability of two claims (A and B) being true [or two events occurring sequentially], written as either p(A&B) or p(AB) can be calculated from either 1) the probability that A occurs (or is true), given B, multiplied by the probability that B occurs or 2) the probability that B occurs, given A, multiplied by the probability that A occurs. And since all probabilities depend on some conditions (usually unstated), then if all these other conditions, C_i, are written explicitly, then compound or sequential probabilities can be calculated from:

$$p(A\&B|C_i) \equiv p(AB|C_i) = p(A|B,C_i) \bullet p(B|C_i) = p(B|A,C_i) \bullet p(A|C_i)$$

That the above rule (or definition or axiom) can be used to calculate the compound (or sequential) probability of any number of claims (or events) can be demonstrated by taking A and/or B, themselves, to be compound or sequential events.

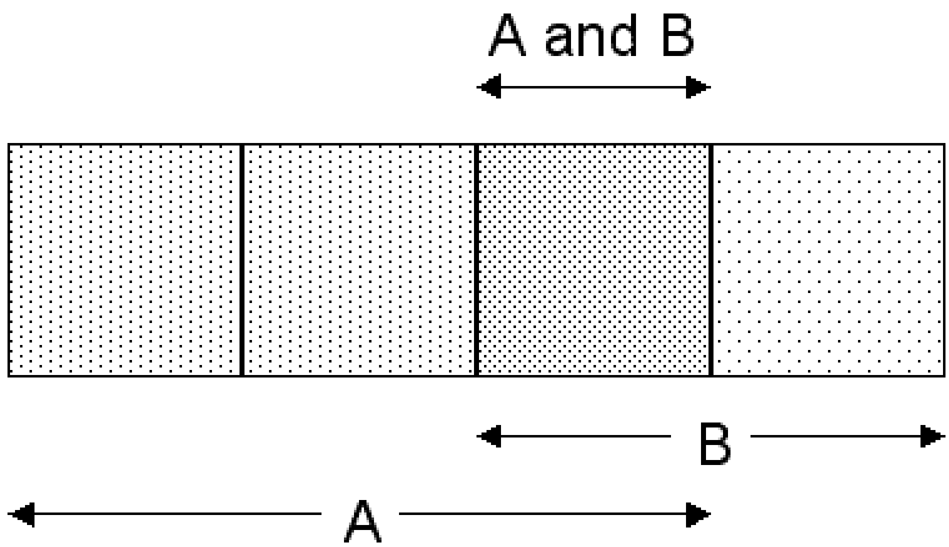
As a first step toward understanding these different types of probabilities, let me suppress explicit identification of the "other conditions", C_i (although they're still to be understood as present). With that, the above "definition" for compound probabilities simplifies to

$$p(A\&B) \equiv p(AB) = p(A|B) \bullet p(B) = p(B|A) \bullet p(A)\ .$$

Next, Dear, please consider the following "Venn diagram" (which Venn created from Euler diagrams of sets and their subsets to display probability relations) – and please continue to consider the diagram until you understand the "answers" shown at the bottom of the figure. For the four "blocks" shown, you might want to interpret "A" to mean "the block is made of an Alloy" and "B" to mean "the block is Blue":[8]

[8] Copied from an article at http://www.statisticalengineering.com/bayes_thinking.htm entitled "Bayesian Thinking" by Charles Annis, P.E.

Venn Diagram illustrating absolute, conditional and joint probabilities.



P(A) = 3/4

P(B) = 2/4

P(A and B) = P(AB) = 1/4

P(A|B) = P(AB) / P(B) = (1/4) / (2/4) = 1/2

P(B|A) = P(AB) / P(A) = (1/4) / (3/4) = 1/3

To try to ensure that you understand the calculated, conditional probabilities given at the bottom of Annis' diagram, maybe the following will help. Suppose that a friend blindfolded you, that he would pull a block out the container, and that your job was to specify relevant probabilities – after he supplied you with partial information.

For example, suppose that, after he confirmed that your blindfold was secure, he pulled a block out of the container, and said: "OK, I have a blue block in my hand; what's the probability that it's an alloy?" You'd need to remember: there were only two blue-blocks in the container and one of those was made of an alloy. So, you'd respond: "The probability that the block in your hand is an alloy, given that it's blue, p(A|B) = 1/2." Good job!

Next, after replacing the blue block in the container, suppose he said: "Ahh, that was just a lucky guess. Try this one. I now have an alloy block in my hand. Tell me: what's the probability that it's blue?" Well, again you'd need to remember: there were three alloy blocks in the container, only one of which was blue. So, you'd respond: "The probability that the block in your hand is blue, given that it's an alloy, p(B|A) = 1/3." Right again – as I trust you'll go back to the previous page and check.

From which I hope you'll notice two points and prepare yourself for a third. The first point I hope you'll conclude is: it's all rather obvious, just "common sense reduced to calculations". Second, notice that by using the calculations shown in the figure, Annis didn't need to work so hard as you did (with your blindfold and your memory) to work out the conditional probabilities: using the formula is much easier! And what I want you to get prepared for is: the same formula, the same method (called Bayes' method), can be used to calculate conditional probabilities for cases when using "just" common sense is much more difficult.

Before showing you what I mean, however, notice a simpler case. Thus, in the case that A and B are independent, then the joint probability of AB reduces to p(AB) = p(A) • p(B), in which I've omitted writing all the C_i conditions (either out of laziness or "following convention"); these conditions, however, are still assumed to be "in force". This simplification for independent events, p(AB) = p(A) • p(B), follows, because if A is independent from B, then p(A|B) = p(A) [and similarly, p(B|A) = p(B)], i.e., the same value for p(A|B) is obtained no matter what B is [and similarly, p(B|A) doesn't depend on A if A and B are independent].

As an example, consider again the case of your worrying that you'd ruin your suede jacket if you got it wet in the rain. In an earlier chapter (**E**, dealing with "Evaluating Endeavors"), I assumed (as a first approximation) that the events were independent, but now, consider the compound event more carefully. You want the probability that you'll get your jacket wet if it rains, say p(Wet). Getting your jacket wet is the result of the compound event "it rains AND you can't find shelter" or, equivalently, "it rains AND you're exposed". Thus (using the abbreviations: W = Wet; R = Rain, Ex = Exposed),

$$p(\text{Wet}) = p(W) = p(\text{Rain and Exposed}) \equiv p(R\&Ex) \equiv p(R\ Ex)\ .$$

In the earlier chapter, I assumed that the events Rain and Exposure were independent and therefore used p(W) = p(R) • p(Ex), but is that realistic? In general,

$$p(W) = p(R\&Ex) = p(R|Ex) \bullet p(Ex) = p(Ex|R) \bullet p(R)\ .$$

Now, if you're walking in the desert (without an umbrella), where Joshua trees are zero help in protecting you from the rain – and it's usually foolish to seek shelter under a (sandstone!) cliff – then not only are the events independent, i.e., if it rains you'll be exposed [so, p(W) = p(R) • p(Ex)], but the probability that you'll be exposed is essentially unity. In the desert, consequently, p(W) simplifies even further to p(R).

In the city, in contrast, it's more likely you can find shelter. But then, can you assume that the two events (rain and exposure) are independent? You might think: "Well, of course they are! I heard on the radio this morning that p(R) = 30%, and although the world sometimes seems to have it in for me (☹), it ain't gonna change the probability of rain if I wear my suede jacket!" That's true enough, but you want not just the probability that it'll rain, but the probability that it'll rain on you – and you have the smarts (it's assumed) to seek shelter if rain threatens. Thus, you want p(W) = p(R&Ex) = p(R|Ex) • p(Ex) = p(Ex|R) • p(R) .

You might be thinking: "Okay, I get the point. So, to get the probability that I'll get my jacket wet I'll use p(W) = p(R&Ex) = p(Ex|R) • p(R), and with p(R) = 30% (from the weather forecast) and my guess that the probability that I'll be exposed given that it's raining, p(Ex|R) = 40%, then that'll give me p(W) = 0.3 • 0.4 = 0.12 = 12%."

That's the way to do it, kid, but don't think that I didn't notice that you shied away from the equivalent way of calculating the probability that you'd get your jacket wet, namely, p(W) = p(R|Ex) • p(Ex) – and if you continue to do that, then you'll never be able to use Bayes' method! "Well," you respond, "of course I didn't use that second expression – because the weather forecast gave me the probability of rain, p(R), but I don't have a clue what p(R|Ex) is! What's the probability of rain, given that I'll be exposed?!"

A part of what you're saying, Dear, is correct – and it's important – namely, the probability that it'll rain, p(R), is an entirely "different animal" from the probability that it'll rain given that you're exposed, p(R|Ex) – except in the desert! The probability that it'll rain when you're exposed depends on your ability to find shelter whenever rain threatens – and on your ability to dash between shelters between showers! For example, after a few years at university in the rain country of the Pacific Northwest, I got high grades for getting p(R|Ex) almost zero, traveling across campus, between classes, through buildings!

For your case of trying to estimate the chance that you'll get your suede jacket wet, however, use the method for which you have information: not p(W) = p(R&Ex) = p(R|Ex) • p(Ex) , but p(W) = p(Ex|R) • p(R), since you have p(R) from the weather forecast and it's easier to guess the probability that you'll be exposed [= 1 – (the probability that you can find shelter)] given that it's raining, p(Ex|R), than it is to estimate the probability that it'll rain given that you're exposed, p(R|Ex).

Fundamentally, however, the two conditional probabilities, p(R|Ex) and p(Ex|R) must be related – because they both depend on your ability to dodge the rain! In fact, obviously the one can be calculated from the other, because

$$p(W) = p(R\&Ex) = p(R|Ex) \bullet p(Ex) = p(Ex|R) \bullet p(R) \,,$$

which, by dividing the last equality by p(Ex), yields

$$p(R|Ex) = p(Ex|R) \bullet p(R) \,/\, p(Ex) \,.$$

For example, "pulling numbers out of the hat", if p(Rain) = 30%, if p(Exposure) = 40%, and if p(Ex|R) = probability that you'll be exposed given that it's raining = 20%, then

$$p(R|Ex) = 20\% \bullet 30\% \,/\, 40\% = 15\% \,,$$

in which I hope you notice that p(R|Ex) = 15% isn't the same as p(Ex|R) = 20% . Meanwhile, I'm not worried that you didn't notice that you just finished using Bayes' method, because that's what I wanted to show you (☺) and it's what I'll now pursue further – for the case of your trying to determine if your friend is cheating.

The essence of Bayes' method is simply this (stated in non-mathematical terms!): if you're having a devil of a time figuring out a particular conditional probability, then stop sweating it; instead, use Bayes' formula to calculate it! For example, in the case of trying to determine if your friend is cheating, suppose that, to date, your friend has won n-times in succession (i.e., you had to pay for all n-lunches). You're worried that she's cheating (C), i.e., not honest (¬H), because if she wasn't cheating, then the probability that she'd win n-times [$p(W_n|\neg C) \equiv p(W_n|H)$ = the probability that she'd win n-times if she were honest (H)] would be $(½)^n$. So, given the evidence that she won n-times, what's the probability that she's cheating, $p(C|W_n)$, or stated differently [assuming that there's only two possibilities, i.e., either that she's cheating or that she's honest, with p(C) + p(¬C) = 1 = p(C) + p(H)], then what's the probability that she's honest, given that she won n-times, $p(H|W_n)$?

Bayes' formula to the rescue! From the definition (or axiom of probability theory – or from common sense), the joint probability that she wins n-times (W_n) AND that she's cheating (C) [or, if you desire, use the description that she's not honest (¬H)] is given by

$$p(W_n \ \&\ C) \equiv p(W_nC) = p(W_n|C) \bullet p(C) = p(C|W_n) \bullet p(W_n) \ .$$

From that, by dividing by $p(W_n)$ [provided that it's different from zero], then you immediately have a method (Bayes' method) for calculating the probability that she's cheating, based on the evidence that she won n-times:

$$p(C|W_n) = p(W_n|C) \bullet p(C) / p(W_n) \ .$$

If you explicitly identify all other "determining factors", D_i , on which all the probabilities might depend (e.g., whether the restaurant burns down, whether you're dumb enough to continue playing what seems to be a rigged game, and so on), then using more general formalism, "Bayes' theorem" would be written as

$$p(\mathcal{H}\,|E,D_i) = p(E|\,\mathcal{H},D_i) \bullet p(\mathcal{H}\,|D_i) \,/\, p(E|D_i)\ ,$$

in which, for the case of under consideration, the Hypothesis, $\mathcal{H}$, is that she's cheating (C) and the Evidence, E, is that she's won n-times (W_n).[9]

As you might imagine, in this form Bayes' theorem appears to be (and is!) extremely powerful, since it provides a method for calculating the probability that any hypothesis is true, based on the evidence – which is obviously a key step in the scientific method. But such generalities aside until a later chapter (**T2**), the question is: How can Bayes' method be used to estimate the probability that your alleged "friend" is cheating? To answer that question, go back to the simpler expression of Bayes' theorem,

$$p(C|W_n) = p(W_n|C) \bullet p(C) \,/\, p(W_n)\ ;$$

if you know how to evaluate all the terms on the right-hand side, then you've "got it made": a simple way to calculate the desired conditional probability.

The first term on the right-hand side, $p(W_n|C)$, i.e., the probability that she wins given that she cheats, seems to be unity. That needn't be so: if she were a smarter cheat (!), then every once-and-a-while, she'd lose – just to keep you in the game! In that case, it could be difficult to determine the strategy she was using, so she doesn't always win (even though she's cheating). But given that she's won all n-times, start with the assumption $p(W_n|C) = 1$, i.e., the probability that she wins given that she's cheating is (apparently) unity.[10]

The next term on the right-hand side of the above equation, seems to be a "show stopper", p(C) – the (unconditional) probability that she's cheating. When you were worried about getting your suede jacket wet in the rain, R, the corresponding term was the probability of rain, p(R), which you got from

---

[9] Sorry again, Dear, but potential confusions in what I wrote include that I'm now using: "W" as an abbreviation for "Win" (not "Wet"), "D_i" for "Determining factors" (whereas earlier I labeled such "Conditions" with "C_i" – but now I'm using "C" for "Cheating"), and finally, at least I used a "script H", $\mathcal{H}$, for "Hypothesis", since I've used "H" for "Honest". From all of which there's a moral: if you're gonna be a "theoretician", ya gotta learn to roll with the symbolism!

[10] If you finally win a toss, say the 11^{th} toss, then for your calculation for the 12^{th} case, you might want to guess that $p(W_n|C) = 10/11 = 0.91$. In general, the value for the probability that she'll win given that she's cheating must be between 1 and 0.5 – although it could drop below 0.5 if (still cheating) she arranges for you to win! Now that's the kind of friend you want to have!

the weather forecast. Who's gonna supply you with an (unconditional) probability that she's cheating, p(C)?

Answer: you! Dear: whereas what you want is, not the (unconditional) probability that she's cheating, p(C), but instead, the probability that she's cheating based on the evidence, $p(C|W_n)$, then for p(C), pull a number out of your hat (it may be a good as the forecast for rain!), and then, see if the above formalism can give you what you really want: not some guess about the probability that she's cheating, but the probability that she's cheating based on evidence!

But maybe don't choose your (initial) guess, about the (unconditional) probability that she's cheating, entirely at random. After all, she has provided other evidence (it's assumed!) that she's your friend; so, maybe start with an initial $p(C) = p_o(C)$, say, of 5% (or 1% or 10% or similar). [By the way, this initial assumption, $p_o(C)$ or, more generally, $p_o(\mathcal{H})$ is usually called "the prior (or the *a priori* = before the fact) probability (that the hypothesis is true)".]

The final term in the above expression is the denominator, $p(W_n)$, known as the total probability. You can evaluate this from

$$p(W_n) = p(W_n|C) \bullet p(C) + p(W_n|H) \bullet p(H),$$

i.e., the probability that she wins n-times on condition that she cheats multiplied by the (unconditional) probability that she's cheating plus the probability that she wins n-times on the condition that she doesn't cheat (i.e., that she's honest, H) multiplied by the (unconditional) probability that she's honest. This you can approximate, in part using your assumed, initial (or *a priori*) probability that she's cheating, $p_o(C) = 1 - p_o(H)$.

Now, put some numerical values in all that. The probability of her winning n-times in a row (if the coin were fair and if she had no other tricks or mirrors up her sleeve) is $p(W_n|H) = (½)^n = 2^{-n}$. The probability of her winning if she's cheating is assumed to be unity: $p(W_n|C) = 1$ (or so close to unity that the difference will be negligible – unless, of course, she would become a smarter cheat and start losing every once and a while!). With those numbers,

$$p(C|W_n) = 1 \bullet p_o(C) / [1 \bullet p_o(C) + 2^{-n} \bullet p_o(H)] .$$

This result states that your estimate for the probability that she's cheating, based on the evidence, depends on your initial assumptions about her honesty. This is similar to the case when you were trying to determine the probability that the Sun would rise each morning, a problem that you "solved" by starting out with one white ball and one black ball in the bag, i.e., you took $p_0(S) = ½$. And it's similar to the problem you encountered in your jury duty, which you "solved" by assuming that, initially, the accused was as innocent as anyone else in the world, $p_o(I) = 10^{-10}$. If, for the problem at hand, you use the initial assumption that there's a 95% chance that your friend is honest, i.e., $p_o(H) = 0.95$ [or $p_o(C) = 0.05$], then the following table (from d'Agostini's CERN report, referenced earlier) shows the resulting estimates for the probabilities that she's a cheat (or that she's honest), as she wins progressively more times, n, i.e., based on the evidence.

n	**$p(C\|W_n)$ %**	**$p(H\|W_n)$ %**
0	5.0	95.0
1	9.5	90.5
2	17.4	82.6
3	29.4	70.6
4	45.7	54.3
5	62.7	37.3
6	77.1	22.9
…	…	…

I trust you notice that, as n increases, those numbers are pretty much what you expected based on your earlier, "ball-park estimates", based on common sense – although they're slightly more "generous". That is, when you did the calculations in your head, you thought "somethin' ain't right" when she won six in a row [for which the unbiased probability is only $(1/2)^6 = 1/64 \simeq 1.6\%$, but the above table suggest that the probability that she's cheating is only 22.9%]. But your "ball-park estimate" didn't take into account the bias that you thought she was honest – maybe wisely so! To illuminate that point, d'Agostini shows the following table illustrating the consequences of different initial assumptions about your friend's honesty:

$p(C|W_n)$ %				
	n = 5	n = 10	n = 15	n = 20
$p_o(C)$				
1%	24	91	99.7	99.99
5%	63	98	99.94	99.998
50%	97	99.90	99.997	99.9999

This second table illustrates two important points. One is that, if you had initially assumed that there was a 50:50 chance that she was cheating (as given in the final line of the above table and as, in fact, you did assume, by ignoring such considerations in your first "ball-park estimate"), then this more complete analysis gives essentially the result obtained when you did the calculation in your head.[11] And a second point to notice in the above table is that, as the number of trials increases (and as you become "broker and broker"!), your initial estimate for how reliable your friend was [i.e., your *a prior* probability, $p_o(C)$] becomes pretty much irrelevant: no mater your original "belief", after 20 trials, you can conclude with 99.99…% confidence that she's not to be trusted![12]

But then, Dear, there's the rest of the story. After doing the calculation using Bayes' method using $p_o(C) = 5\%$ and n = 10, i.e., after finding (based on the evidence) that there's a 98% chance that she's a cheat, you decide to confront her. After the 11th lunch, she reaches into her purse. You suspect that, once again, she's about to pull out a coin. You interrupt her with: "By the way, I'd like to say something." "Oh, so do I," she responded, "but you go first." In all innocence, you respond "Oh, no; please, you go first."

"Well then," she starts, "I just wanted to say that I've become really quite embarrassed that I won all 10 of the previous coin tosses to determine who would pay for the lunch." You squeezed in: "Yes, it seemed quite improbable." She went on: "Well, then I got to thinking how wonderful it is

[11] To confirm that result, see the entry in the table for $p_o(C) = 50\%$ and for n = 5, and recall that $(½)^5$ = 1/32 ≈ 3.1%. So, your ballpark-calculation would be that there's essentially a 97% chance that she's cheating.

[12] A third point, which you might want to think about, is how this example can be used to illustrate the difference between the "standard of proof" in criminal trials [*viz*., "beyond a reasonable doubt" – a concept that's open to doubt (!), maybe meaning less than one chance in 10,000 or one chance in a million or… that the person is innocent] *versus* the standard in civil trials [*viz*., "preponderance of evidence" – a concept that's also unclear, maybe meaning 90% or 99% or 99.9% or… certain!].

to have a friend who would trust me, even when the evidence was against me, and so, to show my appreciation, I bought you a little gift." Whereupon she pulled out from her purse a beautiful wrist watch, that was easily worth ten times what you paid for the lunches – if not a hundred times more. She added "But I interrupted you: what were you going to say." Sheepishly you responded: "Oh, I'm speechless. This is such a wonderful gift. You shouldn't have. But thank you!"

Well, Dear, I won't insult you by concocting an obvious moral of that story, but I do want to call your attention to the most important lesson in all three of the above examples. In fact, this lesson is not only the most important lesson in the above three example, it's one of the most important lessons in life. And the lesson is simply this: not only should you base your decisions on data (aka evidence), but try very hard to make sure you understand what the evidence really "says".

To try to show you what I mean, let me add additional comments on each of the above three examples.

1. In the example of your trying to determine the probability that the Sun would rise each morning, I mentioned the possibility that you encountered a Sun-god worshiper. He was convinced that the Sun god, Sol, rose each morning because he (and fellow worshipers) made appropriate sacrifices. For him, the Sun's daily rising was evidence that Sol was satisfied, but in reality, the Sun's rising each morning was indisputable evidence only that the Sun rose each morning.

2. In the example of your doing jury duty, the scratches on the defendant's face and the match of the defendant's DNA under the victim's fingernails were taken to be strong evidence of guilt. In reality, however, this was evidence, not that the accused was guilty of murder, but only that his face had been scratched by the victim.

3. And in the example of your trying to determine the probability that your friend was a cheat, the evidence that she had won ten tosses in a row was compelling – but based on the evidence in the form of her giving you an expensive watch, you should reconsider the possibility that her luck at winning ten tosses of a coin was evidence only that she had a streak of good luck.

Therefore again, Dear: please try hard to understand what the evidence really "says"; be careful not to impose your interpretation of what it "says". Stated differently, I'd wish you'd not only seriously consider adopting the motto "Show me the data!", but add: "Let the evidence speak for itself!"

AN "UNWINNABLE" APPLICATION OF BAYES' METHOD

Another example, which I mentioned earlier in this chapter, is of the "Ph.D. theoretical physicist" who recently published a book that purports to show how Bayes' method can be used to estimate the probability of God's existence. As was repeated in many news reports and analyses (thousands of which you can find on the internet using search words such as "Bayes' theorem" +"existence of God" +"Stephen Unwin"), Unwin's "objective conclusion" was that the probability that God exists is 2/3 (although, in at least one interview, he added his "personal opinion" that the probability of God's existence was closer to 95%). In reality, however, although he demonstrates that he understood the mechanics of applying Bayes' method (as any theoretical physicist should), he also demonstrates that he doesn't understand (as any experimental physicist should) what evidence means – and he doesn't know how to let evidence speak for itself!

Now, I don't want to use a lot of space to go into this example in detail, but given my accusation of this fellow's incompetence, let me use enough space to outline reasons for my accusation. He purports to use evidence to estimate the probability that the proposition "God exists" is true, i.e., to use evidence to determine p(G|E), in which I'm using the symbol 'G' for 'God' and 'E' for 'Evidence'. As I showed you a few a pages ago, Bayes' theorem for estimating the probability of the truth of any hypothesis, $\mathcal{H}$, is:

$$p(\mathcal{H}\,|E,D_i) = p(E|\,\mathcal{H},D_i) \bullet p(\mathcal{H}\,|D_i) \,/\, p(E|D_i) \,,$$

in which, again, E is evidence, D_i are any other "deciding conditions" – such as the condition that the universe actually exists! Also, recall that, in the above formula, $p(E|D_i)$ must be different from zero – because, in math, dividing anything (even zero) by zero is a "no-no".

If the D_i aren't explicitly identified (but, of course, they're still in effect!), then for the case that the "hypothesis" is "God exists" (although I'm reluctant to call the proposition "God exists" a "hypothesis", since it isn't a succinct summary of a substantial quantity of reliable data – or even of any data!), then the above equation becomes

$$p(G|E) = p(E|G) \bullet p(G) \,/\, p(E) \,,$$

provided, again, p(E) is greater than zero, i.e., $p(E) > 0$. If the only two options are God or not-God i.e., G or (¬G), then the denominator, p(E), is given by

$$p(E) = p(E|G) \cdot p(G) + p(E|\neg G) \cdot p(\neg G) ,$$

in which p(¬G) is used to identify the probability that God doesn't exist (i.e., the probability of "not God"). By definition, $p(G) + p(\neg G) = 1$.

Now, although I don't have Unwin's book in front of me, the "evidence" that he used can be gleaned from a report prepared by Glen Davis. To find this report (which Davis explains is "inspired by Stephen Unwin's book *The Probability of God*") go to http://xastanford.org/?m=200404 and click on "an Excel spreadsheet". There you'll find a list (the usual list!) of "Evidentiary Areas" for "God's existence", namely, those dealing with "morality, evil, suffering, prayers, miracles, religious experiences, origin of the universe, and logical necessity". On the spreadsheet provided, the reader is instructed to choose (for each of these "evidentiary areas") some number between zero and infinity (using unity for "unsure") to indicate one's opinion "whether or not the evidence in that category is more likely a result of God existing or of a universe without God."

Hello?! Are Davis and Unwin totally clueless about what constitutes evidence for testing hypotheses?

Let me outline what I mean, starting with the "evidentiary area" labeled as "miracles". What's the probability that any "miracle" (in the Biblical sense) has ever occurred? There's evidence that such "miracles" have been reported (e.g., in the Bible, the Quran, and the Book of Mormon), but such reports provide evidence for the existence of such reports – not for the existence of miracles! As I've written before, every scientist worth her salt would love to be able to study a miracle, but never have any data become available. So, in the above formalism, what value is to be taken for the probability of the evidence for a miracle?

Let me outline some crazy things that happen when one tries to use Bayes' method in such a case. Denote the probability of evidence for a miracle by $p(E{=}m)$, with m = miracle. According to the above equation,

$$p(E{=}m) = p(E{=}m|G) \cdot p(G) + p(E{=}m|\neg G) \cdot p(\neg G) .$$

Now, although as far as I know, $p(E{=}m) = 0$, i.e., there's no evidence that any miracle had ever occurred, let me give theists the benefit of the doubt and put $p(E{=}m) = 10^{-10}$. Then, what the theists Davis and Unwin want us to conclude is that, when this $p(E{=}m)$ is put into Bayes' theorem, the result is

$$p(G|E{=}m) = p(E{=}m|G) \cdot p(G) \cdot 10^{10} \ ,$$

which certainly seems impressive (i.e., almost no matter what one uses for $p(E{=}m|G)$ and for the *a priori* probability for the existence of God, $p(G)$, the probability for the existence of God based on "the evidence" seems to be huge (because of multiplication by the factor of 10^{10}). That's what "atheists" (better known as "naturalists") have always meant by saying: "Show us a miracle, and we'll believe!"

For example, when so many people were watching their TVs after the first aircraft crashed into the World Trade Center, how about if a giant hand reached down from the sky and caught the second aircraft in mid-air? And then, how about if a second hand snapped its fingers (or whatever), the crash of the first aircraft was undone, and the first aircraft (pre-crash) was caught by this second hand? And reaching still further into this bag of tricks, how about if a resonating voice boldly proclaimed: **"**Thou shalt not kill!**"** Now that would have been a belief-generating miracle!

But meanwhile, those of us who never witnessed a miracle (in the Biblical sense) have concluded (based on the absence of evidence to the contrary) that miracles don't occur; that is, $p(E{=}m) = 0$. Consequently, having learned that dividing by zero is a "no-no", we return to the definition of compound probabilities [before it was divided by p(E) to yield Bayes' theorem, i.e., we return to $p(G|E) \cdot p(E) = p(E|G) \cdot p(G)$], and from

$$p(G|E{=}m) \cdot p(E{=}m) = p(E{=}m|G) \cdot p(G) \ ,$$

we obtain

$$p(G|E{=}m) \cdot \ 0 \ = \ p(E{=}m|G) \cdot p(G) \ ,$$

from which we conclude that $0 = 0$ (refreshing news!) and that either $p(E{=}m|G) = 0$ or $p(G) = 0$, take your pick!

Of course the theists object. They claim: "Just because you atheists say 'there are no miracles' doesn't mean that there aren't any." That is, in

$$p(E) = p(E|G) \cdot p(G) + p(E|\neg G) \cdot p(\neg G) ,$$

theists are willing to let naturalists use $p(E|\neg G) = 0$, but they maintain "the right" to use a $p(E|G)$ greater than zero. But then, look what happens when that's put in Bayes' theorem:

$$p(G|E=m) = \{p(E=m|G) \cdot p(G)\} / \{p(E=m|G) \cdot p(G)\} = 1 ;$$

that is, if you assume miracles occur and God causes them, then (surprise, surprise) you conclude that miracles occur and God causes them. In fact, if you'd prefer to believe in invisible flying pink elephants rather than gods, the above result can be generalized [deleting the other "deciding factor" (D) that miracles are caused by God], to conclude (from once again obtaining p = 1) that if supernatural stuff occurs, then supernatural stuff occurs! What silliness!

Similar silliness occurs for the "evidentiary case" of prayers. No one has yet demonstrated that prayers to God are any more effective than prayers to the Sun god, Sol, or any other imaginary being. Prayers might help "focus one's mind", but so might any method of focusing one's mind. Therefore, if the "evidence" dealing with praying is identified as $E=p$, we should use $p(E=p|G) = p(E=p|\neg G)$. When this (common term) is used is used in $p(E=p)$, it becomes

$$p(E=p) = p(E=p|G) \cdot p(G) + p(E=p|\neg G) \cdot p(\neg G) = p(E=p|G) ,$$

because when the common term is factored out, it leaves $p(G) + p(\neg G) = 1$. Upon substituting this into Bayes' theorem it yields

$$p(G|E=p) = p(E=p|G) \cdot p(G) / p(E=p|G) = p(G) ;$$

that is, the probability that God exists given the evidence of "the efficacy" of praying, $p(G|E=p)$, is the very same as the (*a priori*) probability, initially assumed for God's existence, $p(G)$ – which is a long way of demonstrating the obvious: if the evidence doesn't say anything about the existence of God, then you don't learn anything new about the existence of God!

Similar silliness arises in every "evidentiary area" listed by Davis. In the area of morality, the general idea advanced by theists (as I've outlined in earlier chapters and will show you more in the **M**-chapters, dealing with Morality) is that, whereas humans seem to have some "innate" ability to judge between good and evil, this "moral sense" was imbued by God – in spite of the account in *Genesis* that God didn't want us to know the difference between good and evil; that is, according to the Bible, our "moral sense" is a gift from Satan! Naturalists, in contrast, suggest that such abilities in humans are some combination of nurture and nature, with our "altruistic instinct" (for example) "genetically programmed" by evolution, since for a relatively weak and vulnerable (but, we hope, intelligent!) species such as humans, cooperation among members promoted our survival.

Well, the evidence seems clear enough – humans do seem to have a "moral sense" – which, however, is evidence only that we do seem to have a moral sense! To test the naturalists' hypothesis that our "moral sense" has nothing to do with any god (a hypothesis consistent with the probability of the existence of God being zero), behavioral biologists have tested obvious predictions from the hypothesis that morality is derived from some combination of nature and nurture. One such "obvious prediction" is that, if our moral sense is some combination of nature and nurture, then other "social animals" (such as monkeys, dolphins, elephants, etc.) should display similar behavior (which would similarly promote their survival). And as I'll show you in later chapters, the evidence supporting that hypothesis is strong: clear cases of cooperation, altruism, and even empathy have been described in the scientific literature for literally hundreds of species. Such results thereby support the hypothesis that morality has nothing to do with any god.

So, the obvious questions are: 1) What prediction(s) follows from the hypothesis that our sense of morality was dictated by some God? and 2) How are such predictions to be tested? Is the prediction that, if we don't "believe", then we'll suffer for eternity in Hell? Well, whereas only dead people can perform the test of that prediction, how about if all who advocate it volunteer?!

What other prediction? Obviously the best test of the hypothesis proposing a link between God and morality would be to "turn off" God, to see what our moral sense would then be. And although I think that such a test has already been conducted (namely, it's what now exists!), yet I'd agree that such is "begging the question". But whereas theists argue that if people don't believe in God, then they'll have no moral sense, then I reply that the data don't support that prediction: in the prisons of our society, for example, the proportion of "believers" is higher than in the general population (although I should add that this result probably reflects more the correlation between crime and lack of education – and in turn, lack of education and belief in God are highly correlated).

But at any rate, Dear, I hope you see my point. Evidence about morality is evidence about morality. To link evidence about morality to any hypothesis, specific predictions of such hypotheses must be made, and then a new batch of data is needed to test that prediction. That's been done for the hypothesis that morality is a natural phenomenon, but not for the hypothesis that morality has anything to do with any god.

Similarly for the "evidentiary case" of evil. Although Epicurus might have stumped theists for centuries with his

> Is God willing to prevent evil, but not able? Then he is not omnipotent. Is he able, but not willing? Then he is malevolent. Is he both able and willing? Then whence cometh evil? Is he neither able nor willing? Then why call him God?

yet, clerics have always been a cunning group and managed to concoct a resolution, namely, that God has given us "free will" and we are being tested to see if we'll choose good over evil. In essence, then, the clerics argue that evil is good (i.e., black is white – or *v.v.*!).

But meanwhile, when it comes to applying Bayes' method, the evidence that evil exists is evidence only that evil exists. What evil would exist with or without God existing requires different evidence. And the most obvious suggestion for theists (such as Unwin – what an amazingly appropriate surname!) is that the best way to examine the question (of what evil would exist in the absence of God) is to turn him off – if only, but only, someone could find the on-off switch!

And similarly for every one of the "evidentiary areas" listed by Davis ["morality, evil, suffering, prayers, miracles, religious experiences, origin of the universe, and logical necessity" – as well as any other "evidentiary area" that's concocted (beauty, "intelligent design", whatever)]: either there is no evidence (and dividing by zero is a "no-no") or there is evidence. But when there is evidence, it's critical to let the evidence speak for itself: "Yes, suffering exists; yes, the universe seems to exist; yes, 'religious experiences' seem to occur (especially if mind-warping drugs are used); yes, whatever – but what's God got to do with it?"!

And though I'm sorely tempted to leave it at that, I think that special mention should be made of the two "evidentiary areas" labeled by Davis as "Origin/Nature of the Universe". That the nature of the universe (especially, the Earth) is so amazingly conducive to the existence of life is evidence that it is – as observed by those who observe it! If it wasn't, we wouldn't! And as for "Origin of the Universe", look at the following two contrasting "hypotheses" (or better, "speculations").

From their experiences with "the nature of nature" as revealed, for example, in radioactive decay (i.e., that nature seems to fluctuate, tunneling through energy barriers whose penetrability seems otherwise highly improbable), some naturalists speculate that the universe created itself *via* a symmetry-breaking quantum-like fluctuation in an original vacuum, resulting in the Big Bang. As Edward Tryon said: "I offer the modest proposal that our universe is simply one of those things [that] happen from time to time."

In contrast, theists would say (if they would state their speculations honestly) that total nothingness spontaneously generated not a Big Bang but a great omnipotent, omniscient, omni-whatever being, who then went about the task of creating the universe in six days (which, unfortunately, tired him out considerably, requiring that he use the seventh day for R&R). That's quite an extraordinary claim! As Carl Sagan said: "Extraordinary claims require extraordinary evidence."

The point I'm trying to make can be summarized in the single word "evidentialism". As Peter Forrest stated:[13]

> Here by 'evidentialism' I mean the initially plausible position that a belief is justified only if "it is proportioned to the evidence".

I'll summarize my meaning for 'evidentialism' with the following, which I'd even advocate as a "general principle" for your consideration, a principle that I'm tempted to call

[13] At http://plato.stanford.edu/entries/religion-epistemology/.

THE FIRST RULE OF EVALUATIVE THINKING.

> *Just as you shouldn't build skyscrapers on mud and you shouldn't dwell in a house of cards, you shouldn't accept ideas uncritically: adjust the strengths of your commitments to various "beliefs" to be commensurate with the reliability of relevant evidence. Alternatively: for all your ideas, use all available evidence to evaluate the probability of their validity. That is, "believe" nothing; instead, evaluate probabilities and then go with the idea that has the highest expected value. In a word, evaluate!*

Others have said similar, of course, as I'll now illustrate:

> Believe nothing… merely because you have been told it… or because it is traditional, or because you yourselves have imagined it. Do not believe what your teacher tells you merely out of respect for the teacher. But whatsoever, after due examination and analysis, you find to be conducive to the good, the benefit, the welfare of all beings – that doctrine believe and cling to, and take it as your guide.
> [The Buddha (Siddhartha Gautama), c.500 BCE]

> The foolish reject what they see and not what they think; the wise reject what they think and not what they see. [Huang Po (a Zen master who died in about 850)]

> A wise [person]… proportions his belief to the evidence. [David Hume]

> To believe without evidence and demonstration is an act of ignorance and folly.
> [Volney]

> In religion and politics, people's beliefs and convictions are in almost every case gotten at second-hand, and without examination, from authorities who have not themselves examined the questions at issue but have taken them at second-hand from other non-examiners, whose opinions about them were not worth a brass farthing.
> [Mark Twain]

> The house of delusions is cheap to build but drafty to live in. [A.E. Housman]

> For ages, a deadly conflict has been waged between a few brave men and women of thought and genius upon the one side, and the great ignorant religious mass on the other. This is the war between Science and Faith. The few have appealed to reason, to honor, to law, to freedom, to the known, and to happiness here in this world. The many have appealed to prejudice, to fear, to miracle, to slavery, to the unknown, and to misery hereafter. The few have said "Think"; the many have said "Believe!"
> [Robert Ingersoll]

Faith [is] belief without evidence in what is told by one who speaks without knowledge, of things without parallel. [Ambrose Bierce]

It is wrong always and everywhere for anyone to believe anything on insufficient evidence. [William Kingdon Clifford]

The improver of natural knowledge absolutely refuses to acknowledge authority, as such. For him, skepticism is the highest of duties; blind faith the one unpardonable sin… The foundation of morality is to… give up pretending to believe that for which there is no evidence, and repeating unintelligible propositions about things beyond the possibilities of knowledge. [Thomas Henry Huxley]

Facts do not cease to exist because they are ignored. [Aldous Huxley]

We should be agnostic about those things for which there is no evidence. We should not hold beliefs merely because they gratify our desires for afterlife, immortality, heaven, hell, etc. [Julian Huxley]

What a man believes upon grossly insufficient evidence is an index into his desires – desires of which he himself is often unconscious. If a man is offered a fact which goes against his instincts, he will scrutinize it closely, and unless the evidence is overwhelming, he will refuse to believe it. If, on the other hand, he is offered something which affords a reason for acting in accordance to his instincts, he will accept it even on the slightest evidence. The origin of myths is explained in this way… So long as men are not trained to *withhold judgment in the absence of evidence* [italics added], they will be led astray by cocksure prophets, and it is likely that their leaders will be either ignorant fanatics or dishonest charlatans. To endure uncertainty is difficult, but so are most of the other virtues. [Bertrand Russell]

Credulity is belief in slight evidence, with no evidence, or against evidence. [Tyron Edwards]

In spite of all the yearnings of men, no one can produce a single fact or reason to support the belief in God and in personal immortality. [Clarence Darrow]

Faith is the great cop-out, the great excuse to evade the need to think and evaluate evidence. Faith is belief in spite of, even perhaps because of, the lack of evidence. [Richard Dawkins]

I am an atheist because there is no evidence for the existence of God. That should be all that needs to be said about it: no evidence, no belief. [Dan Barker]

We ought to do what we can towards eradicating the evil habit of believing without regard to evidence. [Richard Robinson]

> The importance of the strength of our conviction is only to provide a proportionately strong incentive to find out if the hypothesis will stand up to critical examination. [Peter B. Medawar]
>
> Conviction is something you need in order to act… But your action needs to be proportional to the depth of evidence that underlies your conviction. [Paul O'Neill]
>
> Don't believe anything. Regard things on a scale of probabilities. The things that seem most absurd, put under "Low Probability", and the things that seem most plausible, you put under "High Probability". Never believe anything. Once you believe anything, you stop thinking about it. The more things you believe, the less mental activity. If you believe something, and have an opinion on every subject, then your brain activity stops entirely, which is clinically considered a sign of death, nowadays in medical practice. So, put things on a scale or probability and never believe or disbelieve anything entirely. [Robert A. Wilson]
>
> Believe nothing with more conviction than the evidence warrants. [Arthur M. Jackson]

In whatever form you prefer this "first rule of evaluative thinking" [and I admit that I prefer the last two (listed above) more than what I wrote in italics at the start of this long paragraph!], I trust you consider the "rule" to be totally obvious.

As an example, Dear, suppose I asked you if you "believe" (think, trust, have confidence) that it'll be a sunny day today. Essentially automatically, I'm sure, you'd *adjust the strengths of your commitments to [such a belief] to be commensurate with the reliability of relevant evidence*. You'd recall the latest weather report you saw or heard, you'd remember yesterday's weather conditions (possibly recalling that "persistence forecasts" are amazingly accurate!), you'd think of the weather you experienced when you were out jogging this morning, and so on, and then probably respond something similar to "Well, I'd give it a xx% chance that it'll rain."

In fact, from that example and from the above quotations, I'd entertain a reformulation of the "first rule of evaluative thinking" such as:

> *For all your "beliefs" (or ideas), use all available evidence to evaluate the probability of their validity.*

I trust you agree, Dear, that you do that essentially all the time. For example, if I asked you if you "believe" (think) that your parents will always try to help you, I expect that your evaluation of all your experiences and expectations would be that you think they'd try to help you at least 99% of the time. If I asked you if you "believe" (think) that so-and-so is a "true friend", then maybe you'd relay that you thought your friend is ~95% reliable. If I asked you if you "believe" that so-and-so is honest, you'd respond… And so on. In each case, your evaluation of the probability of the reliability of your "belief" would be based on all relevant evidence (as derived both from your left-brain analysis and right-brain synthesis of all relevant experiences).

Earlier in this chapter, I sketched how to evaluate such probabilities (or "beliefs"), emphasizing you left-brain's analysis capabilities, trying to show you, as Laplace said, that estimating probabilities is "…nothing but common sense reduced to calculations." I also started to show you how to use evidence to evaluate hypotheses. Now, Dear, it's time to get serious! That is, next I want to address: How can we use data (or evidence) to estimate probabilities [or (equivalently) to reach beliefs] about reasonable, rational, reliable hypotheses for guiding us in how we live, e.g., hypotheses about "the nature of reality" and similar fundamental premisses (or axioms). From such estimates, we should be able to better choose goals and plans and hopes that are realistic, rational, and reliable. As you might expect, it'll take me a long time to answer all such questions – in fact, I'll still be trying to answer them in chapters **Y & Z** – but I'll start on the task in the next chapter, which (I think you'll find evidence to confirm) will still be here waiting for you after you get some exercise!