

How to Analyse Retrodictive Probabilities in Inference to the Best Explanation.

Andrew Holster

"One of the central aims of the philosophy of science is to give a principled account of these judgments and inferences connecting evidence to theory. In the deductive case, this project is well-advanced, thanks to a productive stream of research into the structure of deductive argument that stretches back to antiquity. The same cannot be said for inductive inferences. Although some of the central problems were presented incisively by David Hume in the eighteenth century, our current understanding of inductive reasoning remains remarkably poor, in spite of the intense efforts of numerous epistemologists and philosophers of science." (Peter Lipton, 2000, p.184.)

"Lipton had made a systematic study and defense of a way of reasoning that he called 'inference to the best explanation' (IBE). He argued that this way of reasoning was used commonly in science and everyday life. ... According to Lipton, 'beginning with the evidence available to us' we generally 'infer what would, if true, best explain that evidence.'... in making this argument, Lipton challenged a popular conception of science, namely, that scientists test their theories only by evaluating the accuracy of the predictions that their theories make about future events. But there was a problem with this kind of reasoning, one that Lipton and many philosophers of science had noted. Many regarded 'inference to the best explanation' as little more than a slogan, because no one could say exactly what made an explanation best." (Stephen C. Meyer, 2009)

Lipton and Meyer have a special interest in how we *retrodict past events from present evidence*. This kind of reasoning is the key to *circumstantial evidence for all sorts of existence or occurrence claims*, typically where we have witnesses with memories of observations and possibly some permanent records such as photos, and we have to try to judge *whether we can verify their claims of past events from this combination of presently available evidence*. It is also closely related to the 'problem of retrodiction' in philosophy of physics,¹ a central unsolved problem underlying 'reversibility paradoxes' in physics. I will mention this connection to dispel an important myth that the conventional *philosophy of physical time* has cast over the subject, (the myth of *physical time symmetry*), but the main aim of this paper is to provide a *probability model* that

¹ Watanabe 1955, 1965; Holster 2014.

supports the *inference to the best causal explanation* approach. The subsequent aim is to put it into action with real life case studies, using it to work through evidence. It is planned to follow this theory up with two real-life case studies, for their own interest as well as to illustrate the analytic technique. The first involves reports of the (endemic NZ) *South Island Kokako (SIK)*, recently declared an extinct species (*circa 2002-5*). The second involves reports of (US) *Big Foot (BF)*, the large dark hairy humanoid creature reported in the US (considered a mythological creature in science). This article is motivated not from an overly philosophical point of view, but from a practical point of view, and aims to provide a useful scientific technique to help dissect the debates.

The technique is one of breaking subjective estimates down into smaller parts, but it does not replace the need for researchers to make subjective estimates. In fact, the conclusion is emphasized that *there is no purely objective point of view or judgment of the probabilities involved, there is inherent variability between judgments by different witnesses and analysts, because different people must bring different a priori judgments to the analysis*. There is however a common framework into which all these different judgments can be put and compared. I think Lipton and Meyer are right that the framework should be based on IBE (inference to the best explanation). The technique proposed here provides a quantitative framework to help structure the way we dissect the evidence.

We use this kind of technique of *dissecting evidence* intuitively all the time. In terms of breaking down estimates, consider a question like: *How many NZers will find they have breast cancer in NZ this year?* NZ has about 4 million people. Well, a quick intuitive ‘semi-educated’ guess to scale: about 1,000? At least 100, but 10,000 seems a bit much. (I guess this because I know the road death toll is about 300 and there must be more cases of BC than this; while 10,000 is too large at 30 times the road toll.) But this guess (1,000) is crude, and I make this more accurate by breaking down the cases. Thinking aloud: *BC*

affects predominantly females, so we consider the population of 2 million instead of 4 million. Also it occurs mainly in the second half of life, say from about 30 – 70, so roughly 1 million people mainly at risk per year. Now what would the total rate be like for all women over their lifetimes? I guess about 10%. (I would say: certainly higher than 1%, certainly lower than 50%, but hardly more than 20%? Probably comfortably within 5% to 20%). To get this total rate, we can take 1 million, divide by 40 years, and multiply by 0.1 per year, which is 2,500 cases per year.

My first guess of 1,000 per year was on the right scale then - it would give about 4% total population rate, which is plausible. But I think 2,500 is probably better now.

The aim of this little exercise is simply to illustrate that *breaking the estimates down into smaller parts gives us a better understanding and probably improves our estimates*. Before I broke it down, I didn't know what it would mean for the population scale, now I can see that 1,000 per year translates to about 4% in the female population life time, that is very useful. But I also see I can't really discriminate much within my guess of 5% - 20%, or about 1,000 – 5,000, although I expect it to more around 10%. Nonetheless now I have some quantifiable idea of my *uncertainty*.

Now I just did that exercise out of my head, from my 'background knowledge'. "*It isn't scientific!*" you might complain. "You should go and look up *what the scientists have measured* or study it carefully before presenting anything as a scientific result!" In this case, sure, the information is there, we can look it up.² But if the scientist thinks the information is there for everything, they are wrong. If they think the aim of science is to establish the information for everything they are wrong too. They are thinking of some idealized, sterilized,

² The first statistic on the first website found [[US Breast Cancer Statistics](#)] tells us: "*About 1 in 8 U.S. women (just over 12%) will develop invasive breast cancer over the course of her lifetime.*" NZ is similar demographically. A further search gives the latest NZ figures from the Ministry of Health, which in 2014 go up to 2010: 2008 → 2713; 2009 → 2759; 2010 → 2791. So the 'refined' estimate of 2,500 turned out to be surprisingly close – a bit of a lucky fluke perhaps, but it shows the point that *breaking down the estimates and thinking through their details improves them*.

perfected ‘science’, not the challenge of pioneering scientific explanations. The challenge is to *apply a scientific method*, and estimating quantities *subjectively* is an intrinsic and essential part of that method, not just something that we do temporarily will we gather more ‘scientific information’!³

The probability theory in the next pages is there to justify the concepts, but to *apply the method* we only need a couple of essential concepts: first, that of separating our *a priori* expectations or probabilities from our *conditional* expectations or probabilities, and dissecting these separately. Secondly, iterating evidence from multiple cases, through the (Bayesian) concept that we should modify our *a priori* probabilities in light of evidence, to *a posteriori* probabilities; and use these in turn as the new *a priori* probabilities for evaluating the next evidence.

In the remainder of this section, I briefly explain how this approach underpins the larger idea of *inference to the best explanation* (IBE), as advocated by Lipton and Meyer and other philosophers.

To go back to Lipton, he observes that ‘*our current understanding remains poor in spite of the intense efforts of numerous epistemologists and philosophers of science.*’ What is remarkable to me is that, in all this philosophical literature on the subject, there appears little explicit treatment of the *probability model for causal explanations*, such as given here, to

³ It would be interesting to test this ability. Give people standard estimation problems, get them to write down their best guess after 10 seconds, 1 minute, 2 minutes, 5 minutes, 10 minutes. I think everyone would improve their estimates generally speaking after thinking them over more, but how *much* would people improve? Do they hit a plateau, or keep improving? This could also be designed to test against alternative hypotheses, like the theory of morphic resonance proposed by Rupert Sheldrake, which means information and knowledge is shared, transferred, generated, accessed, through another medium than prototypical causal physical exchanges, or physical memory alone. If *some* people turn out to have more spectacular improvements in estimates than most, they may be using better *estimation techniques*; but alternatively, the results might support an alternative explanation like morphic resonance, *psi* abilities, etc. We want to measure the ‘estimation effect’ sufficiently carefully to *discriminate between these two possibilities* if possible.

accompany the *inference to the best explanation theory*.⁴ ‘Explanation theorists’ in philosophy of science make some very good points; but treatments rarely provide analysis of *probability models for their theories*.

The model here is simple enough, indeed fundamental, the diagram of probabilities is the classic pattern of conditional probabilities, so this is not new. But applying it to the situation of *inference to the best causal explanation*, we add a special condition, viz, *it must be governed by future-directed (or predictive) causal laws*. The interpretative principle in physical terms is that:

- *All the conditional probability laws governing the causal sequence are future-directed, not past directed.*

In the model of inference here, this is reflected in the interpretative principle:

- *Retrodictive conditional probabilities are broken down into a priori probability ratios and predictive conditional probability ratios (defined for alternative explanations).*

The first point above may seem common sense to most people, but to philosophers of physics it may seem ignorant – because they generally believe that “*all the laws of physics are time symmetric or reversible!*” Hence past-directed laws and future directed laws should be equally acceptable. But this view is wrong. This conventional analysis of *time directionality in physics* is now understood to be based on a theoretical mistake, embedded in decades of

⁴ E.g. the Wikipaedia article on *abductive reasoning* states the same ‘inversion’ rule derived below:

“*The probability of infection could then have been conditionally deduced as*
$$p(y||x) = p(x)p(y|x) + p(\bar{x})p(y|\bar{x}),$$
 where “||” denotes conditional deduction.

Unfortunately the required conditionals are usually not directly available to the medical practitioner, but they can be obtained if the base rate of the infection in the population is known.”

However the treatment and interpretation given here is quite different.

philosophical interpretation. But in any case, we are not doing theoretical physics: we are doing real science⁵.

The fact is that *in the real world, which runs to the tune of thermodynamics and real time flow, laws of process are causally structured forwards in time.*⁶ This asymmetry basically means that *the present state can causally control the future state, but the future cannot control the present.* This is the foundation of *law-like support for counterfactual reasoning.*⁷ Now we are not talking about some refined metaphysical nicety here: *it is a simple fact of life that causes go forwards in time, and we reason counterfactually forwards in time.* For instance, a bear walking out in front of a trail camera will *cause it to take a photo*, and we can consider how probable it is that a bear on the scene *will cause the bundle of effects that represent the evidence of this, compared to alternative causes of the same evidence.* The reverse probability makes no intuitive sense as a *cause*: *how probable is it that the effect will cause the cause?* But we have no need to argue about it in any case: you can easily concoct special cases where cause and effect laws are probabilistically interchangeable (e.g. in thermodynamic equilibrium), but our model applies to normal cases of *forwards causation*. This is not meant as a general theory for all explanations, it is meant to get an essential probability-theoretical perspective on the generic IBE theory.

One repeated complaint in the field is that, while we intuitively judge the *power of explanations against each other*, the philosophical theorists fail to provide any good definition or measure of *power*. This probability model does

⁵ The broader problem however is that there is a false treatment generally throughout the philosophy of science of *time-directed inference and time-directed causality and time-directed probability*. This stems from the failure in philosophy of physics to give any realistic account of the directionality of time. See Holster [2014] for more detail on this extensive controversy, which infects the whole of conventional philosophy of physics.

⁶ I do not dismiss the idea that there could be backwards causation, or precognition, clairvoyance, etc – but experimental designs to test those phenomena require special consideration – since they challenge the ordinary causal framework itself – the method here applies to ordinary physical evidence, in our ordinary causal framework.

⁷ See Holster (2014) , Chapter 2.

give a direct measure of the *power of the evidence to support some claim, C, against its alternatives, NOT-C or ~C*. This is the *conditional ratio*. This is the ‘power of the evidence’. But what it also reveals is a second, independent factor, the *a priori ratio*, which is the judgment or expectation we bring to situation *before we see the (new) evidence*. This is independent of the power of the new evidence, but the new evidence should change it, to give the *a posteriori* probabilities. This is the simple Bayesian idea: we modify our beliefs according to the evidence *using a rational probability calculus*.

If there is some expectation by the IBE’ists that their theory conflicts with a Bayesian approach or ‘predictive approach’, this technique shows it doesn’t. Meyer’s comments against over-emphasis on *prediction* are right in my view – but this is a different point to the principle that *we break down retrodictive probabilities into the a priori ratio and the predictive conditional ratio*.

More generally, Meyer and Lipton and similar in the philosophy of science approach have analyzed *qualities of explanations* in various ways that give us insights into how we go about judging explanatory power, and they show it is far more subtle and difficult to make a theory of these judgments than scientists think. The probability model puts this into a simple quantitative framework; *interpreting the qualities of explanations as influencing the underlying probabilities*.

“Lipton showed that a good explanation cites a prior cause, typically an event that is part of the ‘causal history’ of the event in question. Further, Lipton showed that to identify the cause of an event, scientists must identify something within the causal history of the event that accounts for a crucial difference – the difference between what did occur and what we might otherwise have expected. ... A good (or best) explanation cites an event that makes a ‘causal difference’ in outcome.” (Meyer, 2009, p.157.)

1. Intuitive reasoning from evidence to past facts.

The intuitive logic of explanation at its simplest uses *habitualised retrodictive probabilities*. This is schematised:

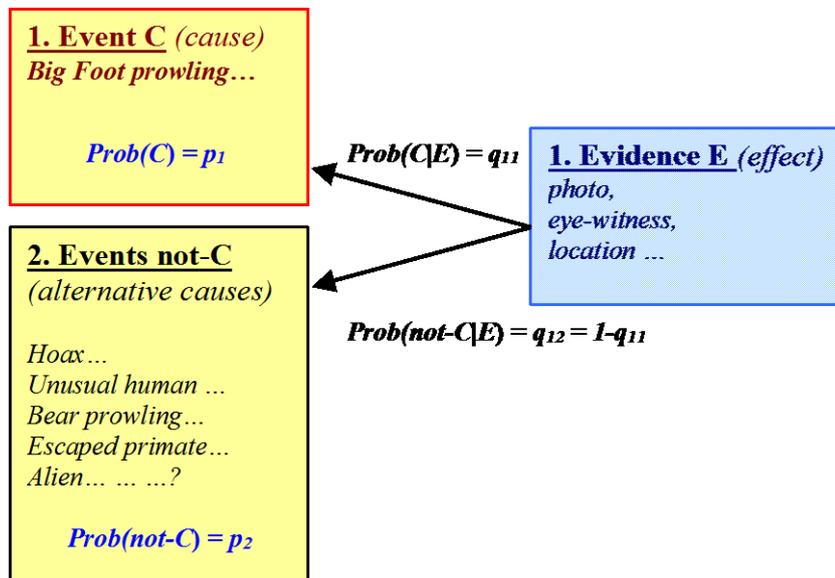


Figure 1.

Evidence E is presented to infer a fact C , e.g. a photograph used to infer a Big Foot sighting; or a sound recording used to infer the survival of South Island Kokako.

When we see a photo of something that looks like a Big Foot, how do we evaluate it as *evidence* for a real Big Foot? Or a photo that looks like anything, for that matter? Suppose you see a photo of your friend, how do you infer it is your friend? Generally:

- From evidence E how do we estimate the probability of claim C ?

Intuitively, we use *habitualised retrodictive probabilities*. This because we are *habitualised to interpreting signs of things by law-like rules*. If we see an animal's foot-print in a creek-bed, we infer a real animal passed by and made it. Our experience is filled with this: *interpreting signs of things that have happened*. Ultimately, the greatest detail is in our memories. A lot is made of the human ability to 'predict the future', but *reading the past is the foundation for reading the present*. In normal decision making we don't *predict the future* in the sense of a rationalised

scientific process, rather we ‘read it’. We read it from the past and present. We explain to ourselves *what various present states-of-affairs mean for the future. We tell ourselves a story about it.* We are very good at this intuitive ability, which it should be emphasised involves counterfactual reasoning, but when we hit difficult cases, where evidence is scarce or ambiguous or circumstantial, etc, or people have different background views that influence their judgements, then we need to turn to a more careful analysis of the principles of evidence if we can.

We are concerned with how to treat the evidence, that: *E implies C.* Intuitive *retrodictive conditional probabilities* are defined as usual:

(1) $p(C/E) = p(E \& C) / p(E) = q_{11}$ [Definition of conditional probability]
 [Bayes Theorem]

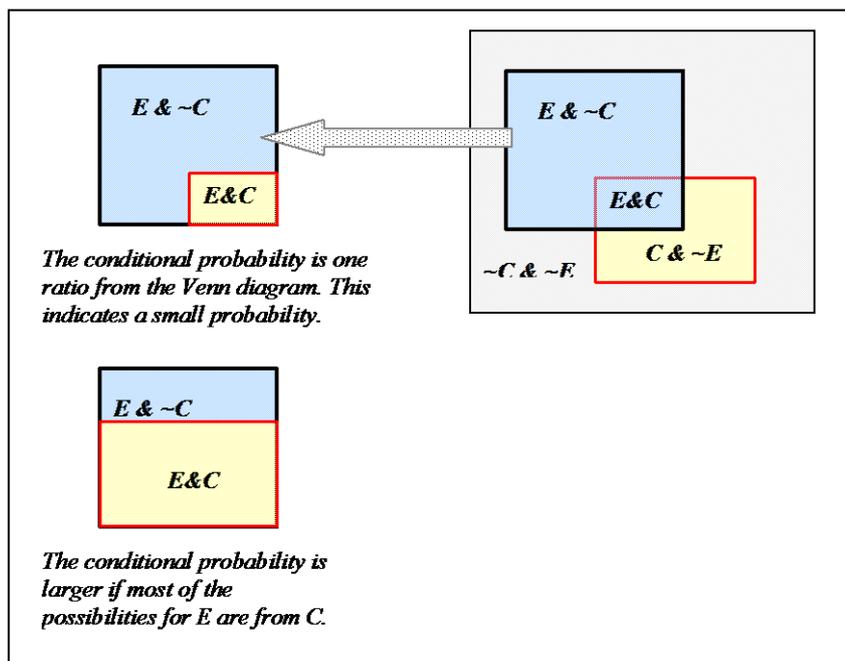


Figure 2.

The idea is *if we know that E (evidence) happened,* then the chance of *C* is the proportion of cases where *E and C* occurred, compared to the total number of cases of *E*. This gives our definition: $p(C/E) = p(E \& C) / p(E)$.

Intuitively, the stronger the evidence *E*, the more alternative possibilities, *~C*, it rules out. E.g. *E* might be a *crater on Earth*, and *C* might be the *a priori* rare event of an *asteroid impact on Earth*. *C* is unlikely if there is another explanation from a more

common cause (say a volcano). But if the evidence is detailed enough, it may rule out volcanos and other causes, and leave us with a rare and unusual event.

If we could determine this *retrodictive conditional probability*, along with the truth of some evidence, *E*, we could infer a probability for the cause, *C*.

But how do we determine $p(C/E)$? In practise, we go through some intuitive reasoning, which involves *how accurate the evidence appears to be – does it really fit the claims?* And just as importantly, *how good are the alternative explanations?* And just as importantly, *what is the preliminary likelihood of C occurring, based on our previous background experience, before we start to evaluate the new evidence, E?* (The *a priori* chance of *C*).

It is clear that the *retrodictive probabilities* are not fundamental or primitive, because we have to judge them and adjust them according to our judgement of all these other probabilities. We want to break down the *retrodictive probabilities* into more basic components, and we now analyse them as resting on *future-directed* or *causal probabilities*.⁸

⁸ Lipton (2000) says:

“Inference to the Best Explanation can be seen as an extension of the idea of ‘self-evidencing’ explanations, where the phenomenon that is explained in turn provides an essential part of the reason for believing the explanation is correct. For example, a star’s speed of recession explains why its characteristic spectrum is red-shifted by a specified amount, but the observed red-shift may be an essential part of the reason the astronomer has for believing that the star is receding at that speed. Self-evidencing explanations exhibit a curious circularity, but this circularity is benign. The recession is used to explain the red-shift and the red-shift is used to confirm the recession, yet the recession hypothesis may be both explanatory and well-supported. According to Inference to the Best Explanation, this is a common situation in science: hypotheses are supported by the very observations they are supposed to explain.”

This seems confused to me. I don’t think there is any such thing as a ‘self-evidencing explanation’ going on in this example. Lipton fails to emphasize the primacy of the *cause* → *effect* relation, and the secondary or derivative nature of the *effect* → *cause* relation.

2. The causal picture.

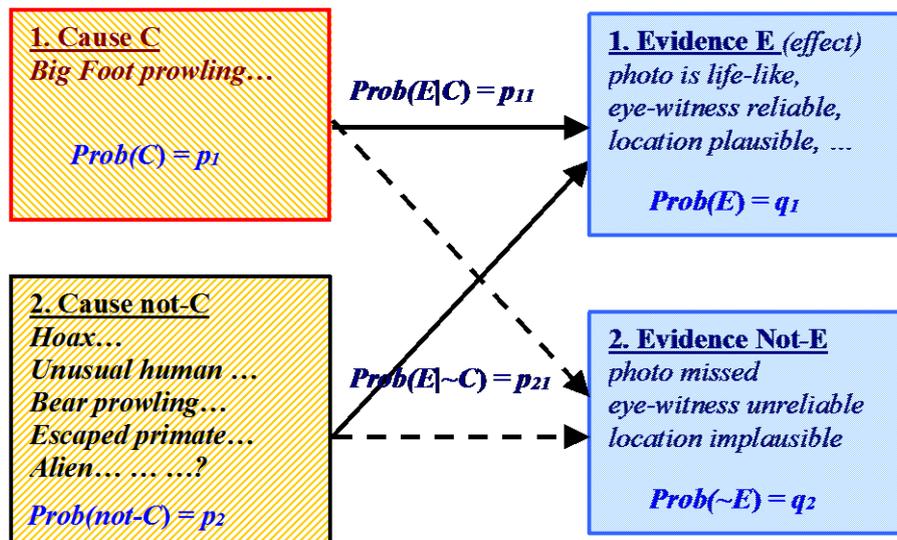


Figure 3.

In the causal picture, we take C to be the *cause* of E , which is its *effect*. The only reason we think the converse (in this naturalistic setting), i.e. that E is *evidence* for the *claim* that C , is because we think that C is a good candidate as a *cause* of E .

We need four probabilities to determine the *retrodictive conditional probabilities*:

- Two *causal conditional probabilities*, p_{11} and p_{21}
- Two *a priori probabilities*, p_1 and p_2

First we define symbols for all the probabilities in the diagram.

$$\begin{aligned}
 (1) \quad & p(C/E) = p(E \& C) / p(E) = q_{11} && \text{[Retro conditional probability]} \\
 & p(C/\sim E) = p(\sim E \& C) / p(\sim E) = q_{21} \\
 & p(\sim C/E) = p(E \& \sim C) / p(E) = q_{12} \\
 & p(\sim C/\sim E) = p(\sim E \& \sim C) / p(\sim E) = q_{22} \\
 (2) \quad & p(E/C) = p(E \& C) / p(C) = p_{11} && \text{[Causal conditional probabilities]} \\
 & p(E/\sim C) = p(E \& \sim C) / p(\sim C) = p_{21} \\
 & p(\sim E/C) = p(\sim E \& C) / p(C) = p_{12} \\
 & p(\sim E/\sim C) = p(\sim E \& \sim C) / p(\sim C) = p_{22} \\
 (3) \quad & p(C) = p_1 && \text{[Notation for a priori probs]}
 \end{aligned}$$

$$p(\sim C) = p_2$$

$$p(E) = q_1$$

$$p(\sim E) = q_2$$

We will take E and $\sim E$ as mutually exclusive and complete alternatives so:

$$(4) \quad p(E \text{ or } \sim E) = p(E) + p(\sim E) = q_1 + q_2 = 1 \quad [\text{Normalised } q_1, q_2]$$

This means that:

$$(5) \quad p_{11} + p_{12} = 1$$

$$p_{21} + p_{22} = 1$$

❖ Proof. $p(E\&C)/p(C) + p(\sim E\&C)/p(C) = (p(E\&C) + p(\sim E\&C)) / p(C) = p(E\&C \text{ or } \sim E\&C) / p(C) = p((E \text{ or } \sim E) \&C) / p(C) = p(C) / p(C) = 1.$

We take C and $\sim C$ here as mutually exclusive and complete alternatives. There will be possibilities we haven't considered as part of $\sim C$, and we return to this point.

$$(6) \quad p(C \text{ or } \sim C) = p(C) + p(\sim C) = p_1 + p_2 = 1 \quad [\text{Normalised } p_1, p_2]$$

$$q_{11} + q_{12} = 1$$

$$q_{21} + q_{22} = 1$$

Note we use *ratios of absolute probabilities*, so we do not have to estimate absolute probabilities of unknown events, only their relative probability compared to common events. We can now define the retrodictive probability, (1), in terms of the causal probabilities, (2).

$$(7) \quad q_1 q_{11} = p(E) p(C|E)$$

$$= p(E) p(C\&E)/p(E) = p(E\&C) = p(C) p(C\&E)/p(C) = p(C) p(E/C)$$

$$= p_1 p_{11}$$

$$(8) \quad q_1 q_{12} = p_2 p_{21}$$

Rearranging gives:

$$(9) \quad q_{11} = p_{11} (p_1 / q_1)$$

$$q_{12} = p_{21} (p_2 / q_1)$$

2.1 Key quantities.

The *retrodition ratio*.

$$(10) \quad \rho = \rho_0 \rho_{1,2} = q_{11}/q_{12} = (p_1/p_2) (p_{11}/p_{21})$$

The *a priori ratio*.

$$(11) \quad \rho_0 = p_1/p_2$$

The *conditional ratio*.

$$(12) \quad \rho_{1,2} = p_{11}/p_{21}$$

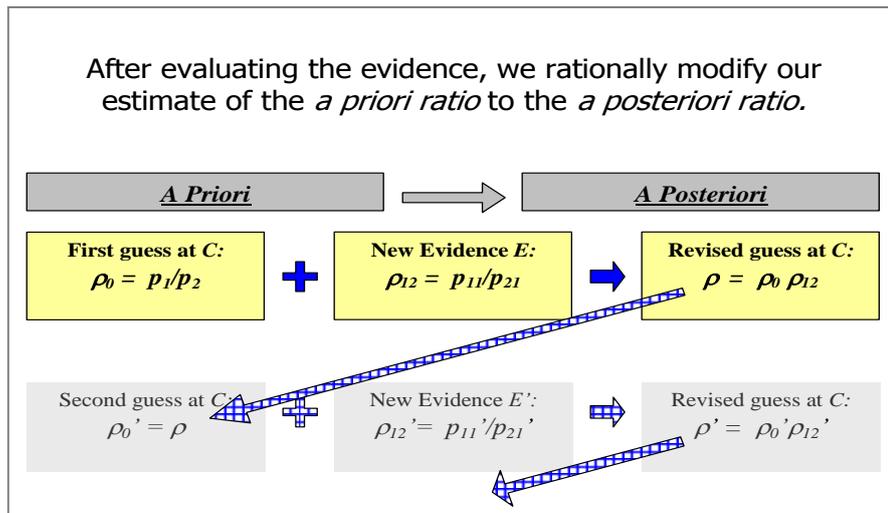


Figure 4

The *retrodition ratio* will replace the initial *a priori* ratio, $\rho_0 = p_1/p_2$, as the new ‘best guess’ for the probability of cause *C* compared to $\sim C$. If we iterate multiple independent evidence for *C*, we get the general expression:

The *multiplied retrodition ratio*.

$$(13) \quad \rho = \rho_0 \rho_{1,2} \dots \rho_{N,2}$$

*Retrodictive probability from predictive ratios.*⁹

$$(14) \quad q_{11} = \rho / (1 + \rho)$$

Retrodictive probability from predictive probabilities.

$$(15) \quad q_{11} = (p_1 p_{11}) / (p_2 p_{21} + p_1 p_{11})$$

⁹ We can solve for q_{11} from ρ using: $q_{11}/q_{12} = q_{11}/(1 - q_{11}) = \rho$.

The simple causal picture with key relationships between predictive and retrodictive probabilities.

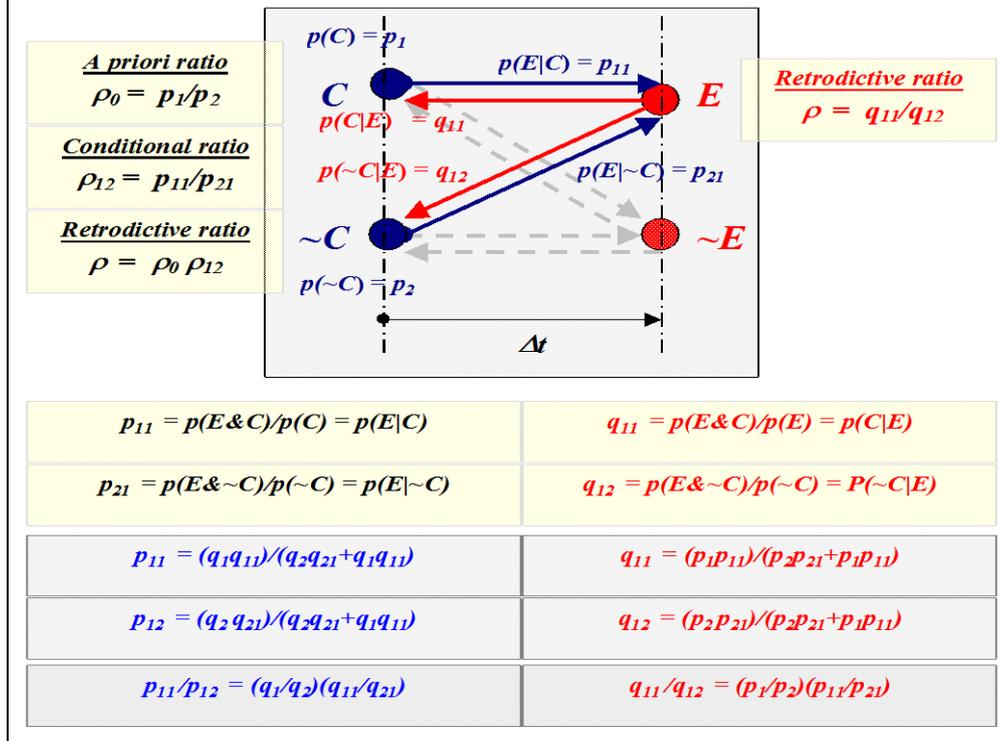


Figure 5

Some basic properties are:

- ρ is the product of two independent ratios.
- ρ is determined by three independent parameters.
- ρ is the key quantity we want to estimate.
- ρ tells us how strongly the event E supports C as the cause, compared to supporting the alternative, $\sim C$.
 - If $\rho = 1$ then C has a 0.5 chance and $\sim C$ has a 0.5 chance.
 - If $\rho = 0.01$ then C has about 0.01 chance.
 - If $\rho = 100$ then C has about 0.99 chance.
- To support C , we want the ratio ρ to be as large as possible.
- To dismiss C , we want the ratio ρ to be as small as possible.

We can now break down disputes directly over retrodiction probabilities to disputes over these components. We can try to distinguish the influence of the *a priori ratio* from the *conditional ratio*. If someone is sure a claim is *false* to begin with, from their *a priori* background beliefs, then they set their *a priori* ratio very low, and *nothing the evidence shows can apparently change their mind*, because the *conditional ratio* can never outweigh their scepticism. If this is their position, then we should disentangle this feature from the evaluation of the evidence. We might then set about examining the evidence for these ratios in turn. And we should emphasise that a system of analysis is needed where evidence is not simply *dismissed and rejected*, it should be used to modify *a priori* ratios in the future.

By working through examples we can see how this gives a model of evidence in a causal probability framework.

2.1 Example: Big Foots.

To illustrate, take a Big Foot report, with a photo and witness, etc. Is it a real Big Foot? We start with our *a priori* estimate of the *probability that Big Foot exist at all*, and consequently could be the plausible cause. This is our first guess at p_1 . Similarly, we can think of our plausible alternative cause, $\sim C$, and guess its likelihood. This is our first guess at p_2 . We can take: p_1/p_2 as our estimate of $\rho_{1/2}$, and *normalise* our initial estimates, see below.

In the problematic case, sceptics about BFs will make a very low estimate of ρ_0 , because they do not credit the chance of BFs surviving in the wild. How small? Should we say: 1/100 or 1/1,000 or 1/1,000,000 ... ? It depends on background knowledge and judgement. Lets start with a 1/1,000th chance, reflecting a ‘sceptical scientific opinion’, but far from an impossibility.

Second, we must estimate the probabilities that *the cause C would have led to the effect E*, and the *alternative cause(s) $\sim C$ would have led to the effect E*. This is where we get into the ‘forensic detail’ of the evidence. If *E* all fits together coherently as an encounter and photograph of a Big Foot, we can give the probability that *C would lead to an effect like E* a value close to 1 – even if we are a sceptic about Big Foots.

- This is a counterfactual inference: *if* a BF walked past the trail camera, *would it* take a photo that looks just like this? (The trail camera, not the BF.)
- The more consistent the photo is with a real BF, the closer this *conditional predictive probability* is to 1.
- If the photo is also consistent with an *alternative cause*, then that could also have a probability close to 1. I.e. the evidence is *ambiguous* between the two.
- The counterfactual probability of a BF causing the photo *E* is low if its content is not very consistent with a real BF.

This is saying the BF event *C* would very likely cause an effect like *E*. But if we subsequently found the BF in the photo was wearing a watch, this would reduce the probability that *the proposed BF event would cause the photographic evidence*, because a real BF would not be wearing a watch. (Or only by the remotest chance). This would increase the chance of the alternative explanation that it was a hoax, since the evidence *E* is more consistent with what we expect from a hoax.

If the eye-witness evidence is very convincing, and there is some photographic or other physical evidence, then this usually helps rule out *alternative explanations too*. Probabilistically, this is because the extra evidence makes it less likely that the hoax or alternative cause *could be elaborate enough to produce all the evidence in such detail*. But the improved evidence must *discriminate between C and ~C*.

		p_1/p_2		
		q_{11}/q_{12}	0.001	0.01
p_{11}/p_{12}	1000	1	10	100
	100	0.1	1	10
	10	0.01	0.1	1

- The cell values give the *conditionalised* chances of *C/~C on the evidence*.
- The *a priori* ratio along the top ranges from 1/1,000th to 1/10th.
- (These are cases where the claim *C* is improbable to begin with.)
- The *conditional ratio* in the column ranges from 10 to 1,000.
- (These are cases where the evidence *E* is *prima facie* well explained by *C*.)

A symmetric chart is below.

retrodictive ratio		p_1/p_2 a priori ratio							
p_{11}/p_{12} conditional ratio		q_{11}/q_{12}	0.001	0.01	0.1	1	10	100	1000
	1000	1000	1	10	100	1000	10000	100000	1000000
	100	100	0.1	1	10	100	1000	10000	100000
	10	10	0.01	0.1	1	10	100	1000	10000
	1	1	0.001	0.01	0.1	1	10	100	1000
	0.1	0.1	0.0001	0.001	0.01	0.1	1	10	100
	0.01	0.01	0.00001	0.0001	0.001	0.01	0.1	1	10
	0.001	0.001	0.000001	0.00001	0.0001	0.001	0.01	0.1	1

- *conditional ratio* = 1 means the evidence E is of no consequence either way, because it is equally likely on either explanation.
- *a priori ratio* = 1 means we give the cause C equal chance with $\sim C$, and the retrodictive ratio is just the conditional ratio.
- *conditional ratio* < 1 means the evidence E counts *against* C .
- *a priori ratio* > 1 means C is the expected cause.
- *a priori ratio* = 10 and *conditional ratio* = 0.01 then *retrodictive ratio* = 0.1. This is a case where C is the normal explanation (in 10 cases to 1); but the evidence E is inconsistent with C , with only $1/100^{\text{th}}$ chance that C causes E ; hence the chance that C is the explanation in this case is only: $10/100 = 1/10^{\text{th}}$.

2.2 Example: Castaway.

The conditional probabilities might both be small in absolute terms, but their *ratios* may be meaningful. E.g. a castaway on a desert island sends a letter in a bottle, C , which is very convincing, but it has only a $1/10^{\text{th}}$ chance of being found (let us say). Hence: $p(E/C) = 0.1$. However, the chance of a similarly convincing letter being produced by alternative means (a hoax) is much smaller, say $1/1,000$, for various reasons – small chance anyone has a motive for such a hoax, and knows the details to put in the letter; etc. Then the conditional ratio is: $1,000/10 = 100$. If this letter in the bottle is discovered, then we estimate a 100 to 1 chance that *the castaway explanation is true*, and we go and look for them.

3. Bayesian statistics and modified *a priori* probabilities.

Suppose we have evidence E that is ‘*weakly convincing*’ for C , let us say with a *conditional ratio*, $p_{11}/p_{21} = 10$. Now if the *a priori ratio*, p_1/p_2 , is estimated to be much smaller again, say $1/1000$, then this only gives us a final chance of about: $10/1000 = 1/100$ that the true explanation of E is C . Scientists would not usually say that this is ‘good evidence for C ’, since it *does not establish C as the most likely explanation*.¹⁰ In scientific convention, people want evidence of C demonstrated with 0.95 or 0.99 probability, before they consider it a ‘scientific proof’. Even if the *a priori ratio*, p_1/p_2 , was $1/10$ (i.e. the cause C is not rare, having a 10% chance), the observation E would only give a resulting 0.5 probability of C in this case. Then we cannot *scientifically confirm it to 0.95 confidence level*.

This attitude that science should *test claims against some predefined standard* is an unfortunate fallacy partly encouraged by statistical theory, reflected in the following false principle.

- ***The Totalitarian Protocol*** tells us to *evaluate each claim of evidence, E , individually; if it is not conclusive at a conventional level of confidence (say >0.95), discard it, and conclude the claim is unproven. Examine the next piece of evidence without being influenced by the previous one.*

This reflects a larger false idea of ‘totalitarian certitude’ in science, a simplistic notion that ‘scientific method’ can be defined like a set of bureaucratic rules, conforming to a rational model devised by a policy analyst. In our previous example, although E is not conclusive, it is obviously evidence *supporting C* , because it raises the *a priori* probability of C by a factor of 10! It is just not ‘conclusive evidence’ by itself. What if we had a number of E ’s? Don’t they add up to more than just one?

If we follow the *Totalitarian Protocol*, we will sequentially discard all our individual pieces of evidence, instead of adding them up as we should. (We see this attitude in practise in the SIK case study). But how do we add up multiple evidence?

¹⁰ E.g. we might have all the symptoms of a rare disease, but because it is so rare, it may be more likely we have some common illness along with some unusual but not so rare symptoms added. I.e. the *a priori probability* of the alternative cause is greater than the power of the evidence for the proposed cause.

We can imagine that we collect two or more pieces of evidence, $E = \{E', E'', \dots\}$, with a single proposed cause, C . We can evaluate it simultaneously or sequentially. If we evaluate it simultaneously, then we have to evaluate the *a priori* probability: $p(C)$ and conditional probabilities: $p(E/C)$, $p(E/\sim C)$. The latter are then:

$$p(E|C) = p((E' \& E'' \& \dots) | C)$$

$$p(E|\sim C) = p((E' \& E'' \& \dots) |\sim C)$$

We assume independence of the elements of evidence, E', E'', \dots . In fact, the selection of these as *elements of evidence* should be chosen so they *are* independent, and: $p(E' \& E'' \& \dots) \approx p(E') * p(E'') * \dots$. Then we have:

$$p(E|C) = p((E' \& E'' \& \dots) | C)$$

$$= p(E'|C) * p(E''|C) * \dots$$

So we get the conditional ratio:

$$p(E|C)/p(E/\sim C) = (p(E'|C) * p(E''|C) * \dots) / (p(E'|\sim C) * p(E''|\sim C) * \dots)$$

$$= p(E'|C)/p(E'|\sim C) * p(E''|C)/p(E''|\sim C) * \dots$$

$$= \rho_{12}' * \rho_{12}'' * \dots$$

I.e. if we have a *series of elements of circumstantial evidence*, E', E'', \dots , for one cause, C , then their separate probabilities *multiply*. E.g. if their conditional ratios are all 10, their combined conditional ratio with N pieces of evidence is 10^N .

E.g. $N = 3$ pieces of separate evidence, E', E'', E''' , means:

$$p(E|C)/p(E/\sim C) = p(E'|C)/p(E'|\sim C) * p(E''|C)/p(E''|\sim C) * p(E'''|C)/p(E'''|\sim C)$$

$$= 10 * 10 * 10$$

$$= 1,000$$

To get the same result by treating the evidence *sequentially*, we have to recursively revise the *a priori probabilities* as we go. That is, we start with the *a priori ratio*: $p(C)/p(\sim C)$, and obtain the first *retrodictive ratio*, from (10):

$$\text{Step 1} \quad \rho' = (p(C)/p(\sim C)) * p(E'|C)/p(E'|\sim C).$$

We then have to use this as our revised *a priori ratio* when we evaluate our next piece of evidence, E'' .

$$\begin{aligned} \text{Step 2} \quad \rho'' &= \rho' * p(E''|C)/p(E''|\sim C) \\ &= (p(C)/p(\sim C)) * p(E'|C)/p(E'|\sim C) * p(E''|C)/p(E''|\sim C). \end{aligned}$$

$$\text{Step 3} \quad \rho''' = \rho'' * p(E'''|C)/p(E'''|\sim C)$$

and so on, giving the same result as before. The result is also invariant with respect to the *order* of evidence, E', E'', \dots . This shows that the procedure is logically consistent in a basic way: we get the same result from the same elements of evidence, independently of the order in which we evaluate them, or whether we take them as one piece of evidence, or separate them into multiple elements.

Note that evidence can help *disprove or falsify or weaken the claim C*, just as much as strengthening it. If some evidence, call it E^* , *weakens C*, then we must have: $p(E^*/C)/p(E^*/\sim C) < 1$. I.e. C makes E^* less likely than the alternative, $\sim C$ does. Evidence E^* strongly falsifies C if: $p(E^*/C)/p(E^*/\sim C) \approx 0$. E.g. suppose we have some positive evidence, E', E'', \dots , but then discover some strong anti-evidence, E^* , that *it was all a hoax*. Now the alternative explanation of being a *hoax* is in $\sim C$, and this means that the probability: $p(E^*/\sim C) \approx 1$. On the other hand, $p(E^*/C) \approx 0$, because the evidence E^* of a hoax (e.g. the photo of BF is seen to have a watch) is highly unlikely if the true cause is a Big Foot.¹¹ The treatment of positive and negative evidence is formally symmetric. We can exchange $E \Leftrightarrow \sim E$ and: $C \Leftrightarrow \sim C$ and the equations are invariant.

¹¹ Although theoretically it is possible. There could be a hoax *and* a genuine BF sighting at once. It could be a double-hoax: a real BF arranged to look like a hoax BF designed to disprove probability theories like this. It *could be*. The probability is not *exactly* zero.)

Modifying the earlier probabilities.

After determining the final result after iterated evidence, we may want to go back and reconsider the individual evidence. For instance, if we started completely sceptical, but the full evidence E', E'', \dots generating: $\rho \rightarrow \rho' \rightarrow \rho'', \dots$ makes us certain of C at the end, we should go back to our individual evidence, and re-evaluate the individual explanations. However the *final result for every piece of evidence* is just the final result for all the evidence: *the a posteriori probability for the explanation C given all the E 's*. This is seen by starting with the strongest available evidence *before* testing a specific causal hypothesis, say C''' . We first include E', E'', E''', \dots , as evidence, modify our *a priori ratio* to (13): $\rho = \rho_0 \rho_{1,2} \dots \rho_{N,2}$.

But this is just multiple evidence for a *single existence claim*, C , and does not deal with multiple evidence generally, and another step is needed.

4. Multiple independent evidence.

Usually we have multiple evidence for *distinct claims of a cause of some type*, e.g. multiple Big Foot reports, or SIK reports, from different times and locations. Here we are not attributing the evidence, E', E'', \dots , to a single causal event, C , but to a series of distinct causal events (in different locations for instance), but of similar types, C', C'', C''', \dots , etc. What they all have in common is that they appeal to a common type of cause: they are all 'Big Foot claims' or 'SIK claims', etc. They are related by this. We will call this general existential claim C , and use C', C'', \dots , to include the specific causal event claims.

To represent this we write each claim in two parts: $C' = C \& C_1$, $C'' = C \& C_2$, etc, as before, where we take C to be the generic claim that *Big Foots exist*, and C_1, C_2, \dots , specify *in addition* details of particular Big Foot encounters, proposed to explain specific evidence, E', E'', \dots , etc.

- $C = \textit{Big Foots exist in the wild}$
- $C_1 = \textit{A BF walked past location } X' \textit{ at time } T', \textit{ and person } A \textit{ took a photo...}$

- $C_2 = A \text{ BF ran past location } X'' \text{ at time } T'' \text{ and trail camera } B \text{ took a photo...}$
- *Etc.*

Note that each C' , C'' , .. , implies the general claim C . We should remember that if we disprove a specific claim, C' , we do not disprove C , only the instance claimed in C' . As a general principle:

$$p(C \ \& \ C') \leq p(C) \quad \text{[Probability theory]}$$

This is because $(C \ \& \ C')$ is more (or equally) specific than C .¹²

- We can now distinguish the *a priori ratio* for the general claim: $p(C)/p(\sim C)$, from the *a priori ratios* for the specific claims: $p(C_1)/p(\sim C_1)$, $p(C_2)/p(\sim C_2)$, ...
- We also want to estimate the conditional probabilities: $p(C_i|C)$, i.e. *a priori conditional probability* of the proposed explanation C_i assuming its 'premise' C is true.

We want the latter to estimate the probabilities that *individual cases are BF encounters*, as opposed to the general claim, C . E.g. if we are generally dismissive of BFs, we will dismiss many specific claims. If we subsequently become convinced of BFs, we should go back and re-evaluate specific claims, and some will turn out to be convincing evidence of a BF encounter, and others may not. This is part of a 'paradigm shift' that people experience when they change their minds about fundamental beliefs, and have to re-adjust a whole explanatory framework in some field.

$$p(C_i|C) = p(C_i \ \& \ C)/p(C) \quad \text{[Definition of conditional probability]}$$

¹² If: $C' \rightarrow C$ (i.e. if C' entails C , as in this case), then: $p(C \ \& \ C') = p(C') \leq p(C)$, and the inequality is true.

The evidence specifically relevant to C_i is E_i . We want to compare this against the probability of C and $\sim C_i$. We can see this in another causal probability diagram, with the assumption that C is true.

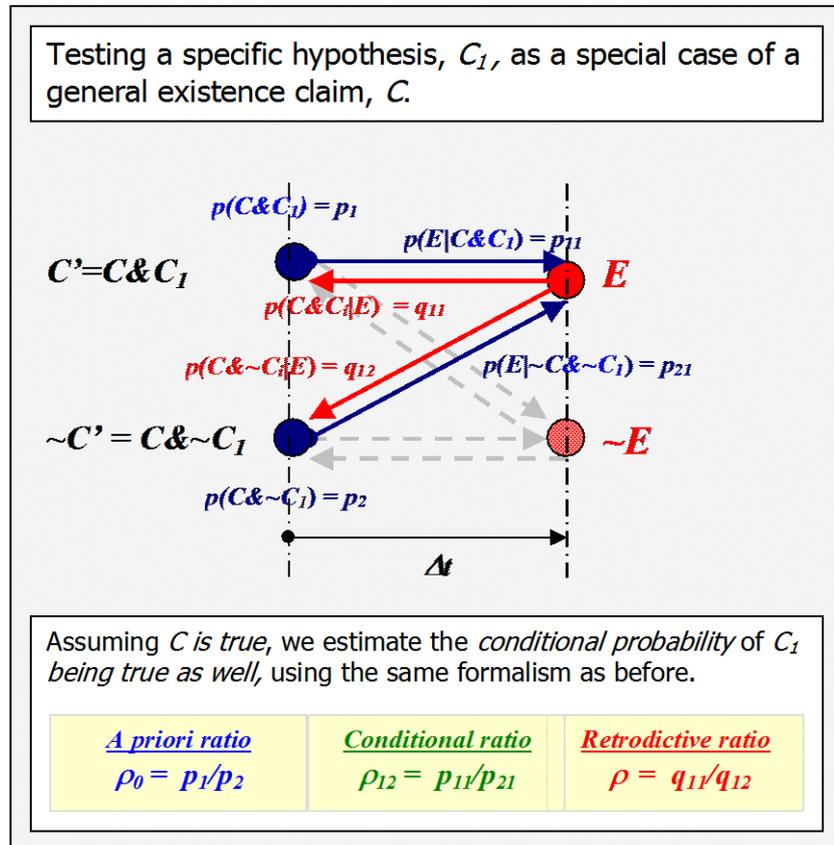


Figure 6

Hence the logic is exactly the same as before, it is all a question of whether we can sensibly estimate the ratios. We now interpret it through some examples.

4.1 Example: Bears and simple confirmation.

E.g. suppose C is the claim that *bears exist*, and C' , C'' , etc. are claims of independent bears in certain locations at certain times. Since C is known, we have: $p(C) = 1$. Beyond that however, each claim C' , C'' , etc. must be taken on its merits. Let us take C' to be a bear in a plausible remote snowy location in the forest mountains, behaving in a normal way, that is difficult for humans or other animals to imitate. C'' is a (apparent) bear in an implausible location, a suburban street, behaving

abnormally, say walking upright and dancing a jig. In both cases we have an eye-witness account, and a slightly blurry photo.

In the first case, we can guess that: $p(C')/p(\sim C') \gg 1$, i.e. the *bear* hypothesis is much more plausible *a priori* than other alternatives that give any likelihood of the effect. E.g. it may be that a well-executed *hoax* could provide similarly convincing evidence, but it is unlikely such a hoax is staged in a remote snowy location where there is a perfectly normal explanation in terms of a real bear.¹³ So we would set the *a priori ratio* of C' much larger than 1, i.e. $p(C')/p(\sim C') \gg 1$.

Note that we avoid estimating absolute probabilities, and use relative probabilities of two choices. Absolute probabilities are ambiguous. For instance, the absolute probability of a person getting a photo of *this bear on this trip at this moment at this exact spot, etc*, will be extremely small, but the probability of getting a photo of *some bear on this trip in this general location* will be much larger. But they will both be vanishingly small anyway – and there is no way we can meaningfully estimate, calculate or compute with such *extremely small* quantities¹⁴. However, whatever background interpretation we take of the bear event, we have to apply it equally to the alternative hypothesis, and it may be jointly comprehended in the (hypothetical) background probability of ‘*getting a photo of some creature like a bear (at location, time etc)*’. (So all we have to add is *the sort of creature it is* to specify the full cause, C' or $\sim C'$.) The smallness of this is what makes the absolute probabilities very small in the first place, but when we take the ratio of absolute probabilities, this cancels out, and we have a moderate-sized probability ratio left that we can sensibly try to estimate. E.g. the absolute probability of getting a photo of a bear may be impossible to estimate, but the relative chances of getting a bear or a wolf may be quite reasonably estimated by an experienced ecologist or hunter.

If the observations also provide *convincing evidence*, it is unlikely that alternative causes could cause the effects (conforming so well to detail), meaning that the *conditional ratio* is much greater than 1, and there is strong overall evidence of a bear.

¹³ How do we *know* it is unlikely? Well, how many times do people see real bears in the woods, compared seeing to *bear hoaxes*? My background tells me the latter is so rare I have never even heard of it, and it would be a difficult, expensive, dangerous and apparently pointless exercise.

¹⁴ Apparently your brain may explode.

In the second case, we can guess that: $p(C'')/p(\sim C'') < 1$, i.e. the *bear* hypothesis is less plausible *a priori* than other alternatives (a hoax; a person in costume). This is because bears are unlikely in the location. People in bear costumes are unusual too, but relatively likely by comparison with *real* bears (in our experience). The *conditional ratio* is also lower than before, because the likelihood of walking and dancing a jig is low if it is a real bear (although it might have been trained), while for a person in a bear costume it is expected. Hence we will conclude that E'' is only weak evidence of a bear in this case.

We will probably conclude that: $C' \& \sim C''$. Now since: $C'' = C \& C_2$, this means: $\sim(C \& C_2)$. But we are certain of C (*that bears exist*), so we infer: $C \& \sim C_2$. Or: “*Bears exist, but this case (the bear dancing in the street) is not a bear.*”

We can write C'' like this:

$$C'' = C \& C_1 \quad \text{and:} \quad \sim C'' = C \& \sim C_1$$

The negation: $\sim C'' = \sim(C \& C_1)$ is logically equivalent to: $(\sim C \text{ or } \sim C_1)$, but if we are certain that C is true, then we can infer: $C \& \sim C_1$. This is only because we are *certain*, in this case, as a background assumption, that *bears exist*, and we are not trying to prove this or question it, and we can leverage off its certainty to simplify the question to whether *this case, C_1 , is of a 'bear'*.

4.2 Example: Big Foots and measures of support or power.

What about the evidence represented by a series of effects, E', E'', \dots , for the general statement C itself? In the Bear example we set $p(C) = 1$ and used *this* to analyse whether a specific report was probably a bear or not. However, in the general case, C is uncertain, i.e. $0 < p(C) < 1$, and we are trying to estimate it. We now go back to the example of the Big Foot, standing for an elusive creature, with a small initial *a priori ratio*: $p(C)/p(\sim C) \ll 1$.

We may assume that some evidence, say evidence: $E_1, E_2, \dots E_N$, supports C , while evidence: $E_{N+1}, E_{N+2}, \dots E_{N+M}$, does not.

Intuitively, evidence E_1 supports C (the claim that *Big Foots exist*) by supporting a specific claim (*the Big Foot C_1 exists*). The evidence E_{N+1} fails to support C because it fails to support a specific claim (*the Big Foot C_{N+1} exists*). However, the latter failure in itself does not *disprove* C of course. It does not disprove C_1, C_2 , etc, because they are about different BFs to C_{N+1} .

MAXIMAL MEASURE OF SUPPORT FOR C.

The most obvious choice is to take *all the evidence that positively supports the existence proposition, C*, and multiply it together, to give the maximum support. Hence we take:

$$\rho_{max} = \rho * (\rho_1 * \rho_3 * \rho_5 * \dots)$$

Where ρ_i are the *conditional ratios*, and all greater than 1. This is the same as multiple evidence for a single cause, C , except *we have dropped all the disconfirming evidence*.

This seems reasonable, but don't the negative cases count for something? And doesn't the number of cases have some influence?

MINIMAL MEASURE OF SUPPORT FOR C.

The *maximum value of the ρ_i 's* can be used as indicating the minimal support for C . This is simply the *strongest support from a single individual piece of evidence*. Obviously this underestimates the evidence, and does not take into account how much evidence there is as a whole. However this kind of statistic is used by the *Totalitarian Protocol*.

MAXIMAL MEASURE ON FILTERED EVIDENCE.

I think in practice we often have set some minimal value for ρ_i , usually at least >10 , as a 'filtering rule' for accepting a piece of submitted evidence, E_i , as supporting

evidence at all.¹⁵ Requiring $\rho_i > 10$ (or more) is practical for a number of reasons. It stops us using a large collection of weak evidence, that we can potentially manufacture, instead of a smaller collection of stronger evidence, that is harder to manufacture. It means there is a higher standard of ‘intuitive convincingness’ required before something is accepted as evidence, so focus on the best evidence. Remember *we are in the realm of using our intuitions to guess probabilities, on a rough scale of 10 or 100 or 1,000 perhaps, not to produce exact numbers*. We can hardly discriminate whether a piece of evidence gives: $\rho_i = \frac{1}{2}$ or $\rho_i = 1$ or $\rho_i = 2$, in most situations. Such low levels of confirmation are too weak to think of as ‘evidence’. We need: $\rho_i > 10$ or more before we can take it realistically as *representing evidence*. Once we accept it, we can treat it realistically, and multiply the ρ_i ’s together.

The choice of such a rule is practical, and depends on the domain we apply it to. In the domain of elusive or cryptic or rare creatures, we have a sceptic’s *a priori ratio* that is very small, say 1 in a million or less, and we have a substantial number of reported encounters of varying quality. Suppose we have 10 observations that all have conditional ratios of 10, then the resulting conditional ratio is $10^{10} = 10$ billion. The probabilities increase exponentially with N . Hence they quickly become astronomical. Once this happens, there is no point continuing to improve them, as the evidence is already conclusive. The real question that re-emerges at that point is whether there is an alternative explanation that has been overlooked.

Another problem with using the simple *maximal measure of support* is that we might make a large number of ‘experiments’, and find enough ‘randomly’ generated positive results to make a false supporting case. Or, more likely, there is a systematic alternative cause ($\sim C$), of some type of low-quality supporting evidence, E^* , which is regularly observed over time, without ever leading to high-quality evidence. By collecting enough examples, we can eventually overcome any low *a priori ratios*, as long as the repeated evidence E^* has a positive (>1) conditional ratio in C . And yet this repeated evidence is not really independent: it is due to one alternative cause.

¹⁵ However it depends what we are doing. If we are searching for a *clue* to solve a baffling problem, we will consider lower grade evidence valuable in our search. But this is not *proving some fact*, it is *searching for something*. Science is not defined by arm-chair definitions of ‘adequate scientific proof’.

Hence *repeated evidence of a certain type* may not count as multiple independent evidence either. E.g. perhaps the most common Big Foot ‘observation’ is *tree-knocking*, loud rapping on trees with a branch, repeated 2 or 3 or more times, usually in the dark. The big foot researchers do tree-knocking themselves, to encourage big foots to answer back. Occasionally they hear other knocks, which could be big foots. But there is usually no corroborating evidence of a BF in these cases because it is at night.¹⁶

The point is: what if there is some completely independent explanation of *tree-knocking*, (say it is a common thing that kids do, a common type of hoax, other big foot hunters, a natural phenomena of wind and trees, or something similar), so the *tree-knocking is not even a real sign of BFs?*¹⁷ How does our model handle this – can it deal with it adequately through probabilities alone, or do we need another rule of evidence to deal with it?

What happens in the probability model is that the lone tree-knocking events *have a low evidence value*, with a *conditional ratio* around 1. This happens if the probability that *a big foot would cause the tree-knocking (if it was there)* is only about the same as the probability that *some alternative cause (kids playing; other big-footers doing tree-knocking; hoaxers; unknown animal behaviour; ...) would cause tree-knocking (if they were there)*. Now in the case of a hoaxer (or some other causes), it is probable they *will* do tree-knocking, just as much as a BF would.

We also do not yet know whether it is true that *BFs do cause tree-knocking*. Or more exactly, whether *if BFs do exist they do tree-knocking*. We know that *if they do exist* they will cause *photographic images* and *visual sightings* and *footprints* and *vocalisations*, in the right conditions. But tree-knocking from this point of view is not direct evidence by itself. Anyway, what we should think is: *even given we accept C*,

¹⁶ They use thermal cameras, but the BFs are by assumption expert at hiding behind trees. Why hide at night when people can’t see you? The BFs have good night vision and may realise people can’t see at night, but they might assume that some people have night vision too. It may be intuitive to follow their customary behaviour of concealing themselves from humans, day or night.

¹⁷ A similar case is ‘*leaf cutting*’, suspected to be connected to SIK, but not certain. Combined with other evidence it is a significant pattern, but by itself it is difficult to take as evidence of existence at all. The fact is that to *establish the existence claim (for BFs or SIK)* we need more robust, direct evidence. Proponents would agree.

and BFs do exist, how much confidence would we have from a single tree-knocking event that it is a BF? As opposed to an alternative cause? It is not strong enough evidence intuitively to give any confidence. (However it may strengthen our confidence in a BF when combined with other observations of the same BF).

Hence this problem of ‘selectively duplicating low-value evidence’ might be dealt with naturally: when the evidence really *is* of low value, it does correspondingly little to substantiate the claims probabilistically. Nonetheless we need to consider more conservative measures.¹⁸

5. The Witness Effect.

Here I note some features of Witnesses.

- Being Witness to an event has a far more powerful effect on belief than second-hand or Analytic evidence.
- Being Witness to a reality can override almost any *a priori scepticism or belief*.
- The Analyst perspective systematically underrates the Witness evidence.
- The Witness systematically overrates their own experience as evidence.

Events are witnessed by human beings, and the treatment of witnesses is very interesting. The Witness best knows the truth of their evidence, and in many cases of Big Foot and SIK sightings, witnesses are adamant about the phenomena. Some witnesses (in relation to all sorts of phenomena) report experiences that must be totally convincing if you experienced them yourself. These are cases where it is clear there is no deception by the witness. From the Witness point of view, the conditional probability ratio becomes very large, equivalent to certainty, because it must outweigh any *a priori* ratio.

Now the fact that our *perceptions* are so strongly convincing makes them to some extent subjective – they give us a *visceral* feeling of certainty, which is beyond objective certainty (especially in a world of advanced human trickery and artifice).

¹⁸ A more conservative measure is: $\rho_{max}/N = \rho * (\rho_1 * \rho_2 * \rho_3 * \dots * \rho_N)/N$. I.e. divide the *maximal measure* by the number of positive tests, *N*. But this doesn't solve the problem.

This is instinctual, consistent with our usual sense experience, which we interpret instantaneously, and use to jump to fast conclusions. In this, we behave as if our beliefs are conclusive – and psychology shows that we readily construct images to fill in the blanks, i.e. we *project certainty on our perceptions that is not there*. Entertaining too many counterfactual contingencies or *probabilities* slows our mental processing and radioactivity.

We should remember that we trust our senses with our lives. If we see a car drive pass on the road, for instance, we interpret it with certainty, and we *do not walk out in front of it*. If we see a giraffe walk down the road, we will check our dream-state, perhaps review any recent hallucinogenic drug use, but finding ourselves befuddled and vaguely irrational as usual, i.e. in our normal state of mind, if we continue to see the giraffe, we will take it for real. If someone asked us the chance of it happening before we saw it, we might say: *one in a million. No one in a billion. No one in a billion trillion. There just aren't giraffe's walking around the neighbourhood. Ever!* If we saw a giraffe walking around our neighbourhood, however, we would quickly overcome this *a priori* probability, accept its reality, and realise that there is an explanation – it escaped from a zoo or someone let it go for a prank or something like that.

This is because although the giraffe event is an *a priori very unlikely event*, it is a very possible event, in that it is quite within human control to cause it. This means it is *a posteriori* a 'mundane' event: once we confirm it, it fits into our known universe.

Seeing a Big Foot is similar up to a point. *If we saw a Big Foot in a realistic setting*, and we got a good close look at it, just like with the giraffe, we would quickly overcome our low initial *a priori* probability, and accept its reality. However the *a posteriori* explanation – *where did the Big Foot come from?* - is not mundane, it is an *existential disruption* to our world view (if we were previously sceptics, that is). It is beyond human control, as far as we know, to cause a Big Foot to appear. Nonetheless there is a sufficiently believable *ecological story* to support the possibility of Big Foot survival, and accept it into our naturalistic universe as a species. Indeed, practically everyone who testifies to a close encounter with a Big Foot takes on this *naturalistic realist view*, as an animal with an ecology.

First-person encounters with *aliens*, to take a further classic example, have a similar rationalisation in the probability framework to Big Foots – and there are people who are conclusive about their first-hand experience – but the *a posteriori* explanation in this case is more *existentially disruptive* again. *If there are aliens abducting people and flying UFOs around the planet, most of us have a huge change of world-view coming up.* But part of our analytic problem is to separate our subjective sense of unreality about such claims (until we experience them) from the rational calculation of the probabilities.

If there are good witness accounts, then for the witness the conditional probability ratio becomes conclusive: say *billions or trillions to one* to over-ride previous strong scepticism of say *millions to one*.¹⁹ But if we have a few good witnesses like this, why don't we take their experience as conclusive? This is a very good question. In fact, to some extent we do. With ordinary crimes, for instance a murder, one witness is taken seriously, two witnesses are very strong, three are practically conclusive. Same with all sorts of things, including sightings of rare but known animals. If there are *three* reliable witnesses to something, people will believe them. Extra numbers get rid of some of the 'subjective element' that we associate with witness.

However the very *visceral feeling of certainty* that we feel as a first-person witness cannot be reproduced with the same strength in the Analyst. Otherwise telling people stories about things would be as powerful as experiencing them first hand. We place a far greater belief in our own *perceptions* than anyone else can place in our *accounts of them*.

In addition, another possibility that ultimately *always* remains is a 'hoax', played by the witnesses, or played upon the witnesses; extending to large-scale conspiracy theories in some cases. Having multiple witnesses does not particularly help in detecting a convincing hoax, it fools everyone equally. We subjectively underestimate how easily *we* can be deceived (although we see it with other people). This is evident

¹⁹ What happens when lets say trillions to 1 chance meets a quadrillion to one chance in your brain? Does it explode?

from conjuring tricks, where we *know* it is a trick, but we *still can't believe our eyes*. Until we learn how the trick is done.

The last explanation of Witness experiences if all else fails is *madness or hallucination*. “You must have been hallucinating.” “What did you smoke before you saw it?” “You’re mad.” “You’re weird”. “Obsessive ideas”. Etc.

With *unconfirmed creatures*, like Big Foots or Aliens, people are markedly reluctant to fully *believe* first-person accounts. Is this objective? Is it because the *a priori probabilities* of these ‘unknown creatures’ are really so much lower than other rare events? Or is it because we systematically distrust witnesses because of subjective vulnerability to hoaxes, etc? Or is it primarily psychological, because *they force an unwanted existential disruption in our world-view*? Or because we are afraid of being associated with ‘madness’?

The *existential disruption in our world-view* is surely a strong factor, and explains the phenomena of *anger* and *social ridicule* for witnesses reporting strange experiences, strange ideas, etc. People are far too opinionated for their beliefs to reflect rationally calculated probabilities, and I don’t care what you say about it! Everyone has noticed this in others. *Other people cling to their paradigms*, and rationalise events to fit their beliefs. But perhaps this rational too? Perhaps it reflects a genuinely lower probability for the disruptive events, than for mundane events? We should hope this is true in general – at least if we want our ‘common sense beliefs’ to be generally accurate. But it cannot be uniformly true – especially with theoretical disputes. And in controversial arguments *between paradigms*, the *a priori* ratio is precisely what is under question, and it cannot be used as evidence.

6. The Sceptic's fallacy.

The ideological Sceptics²⁰ do not want such 'alternative phenomena' investigated, and do not want 'pioneering science' to succeed. There is an interesting interview by Scientific American journalist Horgan of Rupert Sheldrake, a long-standing target of mockery by Sceptics.²¹

Horgan: *If you had the skeptic Michael Shermer (who critiqued morphic resonance in 2005) in front of you right now, what would you say to him?*

Sheldrake: *I would invite him to have a debate about the existence of telepathy and other psychic phenomena. In 2003, in relation to my research on the sense of being stared at and on telepathy, he asserted in USA Today that "The events Sheldrake describes don't require a theory, and are perfectly explicable by normal means." I emailed him to ask what his normal explanations were. He was unable to provide them, and confessed that he had not actually read the evidence. I challenged him to a debate. He accepted, but unfortunately he was so busy being a professional skeptic that he could not find time to look at the data. He has often claimed that "Skepticism is a method not a position." Taking part in this long-delayed debate would provide an opportunity to put his principles into practice. (Horgan, 2014)*

²⁰ The modern creed of 'professional Sceptic' is quite distinct from the tradition of *scientific scepticism*. The latter involves an attitude of *suspension of judgement about claims* until the evidence is carefully evaluated – and specifically compared against 'scientific background beliefs'. 'Professional Scepticism' is an ideological position of *attacking claims and concepts that contradict the Sceptics' favoured materialist metaphysics*. Sceptics only attack specific types of claims – they show little scepticism about 'conventional scientific theories' when these are wrong, inconsistent or untestable.

²¹ Wikipedia, "Rupert Sheldrake", 2014. "*Members of the scientific community who have looked at morphic resonance have characterised Sheldrake's claims as being pseudoscience.*" This neglects to say that other members of the science community take Sheldrake's ideas seriously, and the theory of *morphic resonance* has explanatory power and empirical implications. There is no doubt in my mind that it is a 'scientific theory', and it is *also* an 'alternative metaphysical concept' of causality. (This does not make it 'unscientific': quantum mechanics famously introduced an 'alternative metaphysical concept of causality' too). It contradicts the scientist's *a priori* expectations (or *metaphysical paradigm* about how causation works), but that does not stop it being a *scientific theory*. And this is very much the point: we should not just let "some scientists" *a priori* expectations decide the fate of a theory – it should be tested on its evidence and explanatory power. Sceptics who act outraged at the concept of 'morphic resonance' and make its fate a matter of their own purely subjective judgement are themselves failing to act scientifically.

We should reject the central hypocrisy that Sceptics indulge in, illustrated here, of giving *hand-waving explanations based on a broad theoretical paradigm*, instead of facing the challenge of giving real explanations. It is worth repeating: *“He was unable to provide them, and confessed that he had not actually read the evidence.”*

This is entirely typical. The sceptic claims there is a ‘perfectly normal explanation’ for some puzzling phenomena, for instance psi, but is then unable to specify what the explanation is. They will say *“the psi phenomena is due to some combination of fraud, wishful thinking, bad observation, psychosis, confusion, coincidence, subliminal perception, non-scientific method, ... or something else I haven’t thought of yet.”* When challenged as to the detail, they say: *“I don’t exactly what the explanation is in detail, but I know that there is some normal explanation of this kind – because that is how physics works.”*

We can see this is a ‘hand-waving’ explanation – it doesn’t explain any detail, it just claims that *there is a certain framework for the explanation*. That ‘vagueness’ in itself doesn’t necessarily mean it is wrong though. There is an important place for such ‘framework explanations’, that locate the ‘conceptual frame of reference’ for a more detailed explanation. What about when we try to explain some physical phenomenon, like temperature for instance? “Why is the water hot?” Knowing the atomic theory of heat, we might say: *temperature is the kinetic energy of molecules or atoms bouncing around at the atomic level. The water is hot because it has lots of heat energy.* Now is this a good explanation? Well it sounds all right doesn’t it? Isn’t it true? “*Perhaps*”, we might object, *“but what does it explain? How exactly does kinetic energy translate into temperature? How does it work with water? You have given no specific detailed explanation at all! It’s just hand-waving!”*

Now this is true too – but there are mitigating factors. First, you or I may not know much detail of the atomic theory personally, but we are not relying on our own knowledge, we are depending on our knowledge that some scientists have worked out the theory, and this is a famous result. We have seen it in school some time. We know that this is the kind of explanation that is generally accepted for the phenomenon of temperature.

In addition, it actually *does* tell us a lot: it correctly identifies the connection between molecular kinetic energy and heat and temperature. It doesn't explain each of these in turn, but it correctly identifies them as the components of the explanation. And *explanations don't and can't explain all their claims in turn*. They have to explain *in terms of something*. If you don't know what *molecules* or *kinetic energy* are then you won't understand the explanation, but what it tells you then is that *if you do want to understand the explanation, you need to understand what these things are*. And the previous point means that *you are confident that if you look into molecules and kinetic energy and heat carefully, you will find scientists have explained temperature quite precisely in these terms*.

I think this is how stating a kind of 'framework theory' represents an explanation, and can be a good explanation, at a simple level. But does the sceptic's explanation of *psi* work like this? Well first they can say: "*I am telling you that the high-level explanation of psi involves physical particles, the four known forces, and ordinary causality*." How is this different to telling us that "*temperature is molecular kinetic energy?*" It has a bit of physical mumbo-jumbo in it. But it is not an *explanation*: it is an assurance the sceptic gives about *anything*, and the same proposition cannot be the explanation of *everything*. It is their *a priori* position, and represents their *metaphysics*.

Now this level is what the argument is about in the first place. The *psi* realist says: *look, this phenomena cannot be plausibly explained conventionally, it points to something else*. If the sceptic wants to answer this convincingly, they should find a specific conventional causal explanation. If they can't, they may draw on their previous experience, and say: *I don't know in this case yet, but in the past most cases have been explained*. But now they are on a slippery slope. First, how many cases in the past *remain unexplained*? If *all cases* are explained, that is more convincing. But if there are some cases explained as frauds and hoaxes, perceptual anomalies and hallucinations, some by madness, some by strange weather phenomena, then that is normal and expected, it is human behaviour, but how does that affect the remaining *unexplained cases*? It now depends on *how the sceptic rates the a priori ratio and the conditional ratio*.

However the sceptic explicitly *does not engage in any of this* – their position is that *the a priori probability is so low that it is impossible, and not a permissible science question*. This is where they are deeply confused: *science does not work like that*.

The ideological sceptic might end up being right – but only by accident, and riding rough-shod over the principles of evidence. Their position is anti-scientific. It suppresses the very problem that the psi phenomena poses, that *it is very difficult to find any detailed ‘normal explanation’ that actually works*.

The sceptic’s *position* that “*there is always a normal explanation for everything, we just haven’t found out the details yet*” is anti-scientific. If we applied it to the history of any science, we would find the sceptic opposing the creative scientist who makes new discoveries at every step! (In fact this is exactly what happens in normal science).

“[Shermer] has often claimed that “Skepticism is a method not a position.” Taking part in this long-delayed debate would provide an opportunity to put his principles into practice.” (Sheldrake, 2014).

I conclude that ‘Scepticism’ is *not a method*, it quite obviously is an ideological *position*. It is pseudo-moralistic. It holds that *our current scientific belief paradigm is right and Rational people should believe it and defend it*. Sceptics have to make a special case for *present* science: “*Oh yes, historically science got things wrong, when it was amateurish and only developing, but we have now perfected science, and we know now it is right. In particular, it shows that the fundamental nature of the world is materialist, reductionist, with only local physical causal mechanisms. I know there must be a ‘normal explanation’ for psi even if I don’t know the detailed explanation, because I know this paradigm of existence itself has been established by science generally. It overrides any minor examples you might raise.*”

The analytic technique in this paper *is a method*, and this method is completely at odds with the sceptic’s *position*.

7. References

Holster, A.T. 2003. "The criterion for time symmetry of probabilistic theories and the reversibility of quantum mechanics", *New Journal of Physics*, (www.njp.org) <http://stacks.iop.org/1367-2630/5/130>. (Oct. 2003).

Holster, A. T. 2014. "Principles of physical time directionality and fallacies of the conventional philosophy." 30/7/2014. [<http://philpapers.org/rec/HOLPOP-2>]

Horgan, John. 2014 "Scientific Heretic Rupert Sheldrake on Morphic Fields, Psychic Dogs and Other Mysteries". <http://blogs.scientificamerican.com/cross-check/2014/07/14/scientific-heretic-rupert-sheldrake-on-morphic-fields-psychic-dogs-and-other-mysteries/>

Newton-Smith W.H. (ed) 2000. *A Companion to the Philosophy of Science* (Blackwell, 2000) 184-193.

Lipton, Peter. 2000. "Inference to the Best Explanation." In Newton-Smith, 2000, pp. 184-193.

Lipton, Peter. 1991. *Inference to the Best Explanation*. 1991. Routledge.

Meyer, Stephen C. 2009. *Signature in the Cell: DNA and the Evidence for Intelligent Design*. HarperCollins.

Sheldrake, Rupert. 1988. *The Presence of the Past: Morphic Resonance and the Habits of Nature*. Times Books.

US Breast Cancer Statistics (web page)

http://www.breastcancer.org/symptoms/understand_bc/statistics?gclid=Cj0KEQjwveufBRDlsNb3kb-twMIBeiQASNH0xrR7PVu9yfpzULqU9S11yFbBIJfEyMSTaNtLrpjOKjQaAvQF8P8HAQ

Watanabe, Satoshi. 1955. "Symmetry of Physical Laws. Part 3. Prediction and Retrodiction." *Rev.Mod.Phys.* 27.

Watanabe, Satoshi. 1965. "Conditional Probability in Physics" *Suppl.Prog.Theor.Phys.* (Kyoto) Extra Number, pp. 135-167.

Wikipedia "Abductive Reasoning", 2014.

http://en.wikipedia.org/wiki/Abductive_reasoning

Wikipedia "Rupert Sheldrake", 2014. http://en.wikipedia.org/wiki/Rupert_Sheldrake