[11]. S.F.Gull and J.Skilling, "The Maximum Entropy Method", in *Indirect Imaging* ed. J.A.Roberts, Cambridge Univ. Press, pp. 267–279 (1984).
[12]. J.E.Shore and R.W.Johnson, "Axiomatic Derivation of Maximum Entropy and the Principle of Minimum Cross-Entropy", *IEEE Trans. Information Theory* **IT–26** 26–37 (1980); and**IT–29** 942–943 (1983).
[13]. L.Boltzmann, "Über die beziehung ...", *Wiener Berichte* **76**, 373–435, (1877).
[14]. C.E.Shannon, "A Mathematical Theory of Communication", *Bell System Tech. J.* **27**, 379–623, (1948).
[15]. D.Deutsch, "Quantum theory, the Church-Turing principle and the universal quantum computer", *Proc. Roy. Soc. Lond. Ser A* **400**, 97–117 (1985).
[16]. C.A.Fuchs, "Quantum Mechanics as Quantum Information (and only a little more)", http://www.arxiv.org/abs/quant-ph/205039 (2002).
[17]. C.M.Caves, C.A.Fuchs, R.Schack, "Unknown quantum states: the quantum de Finetti representation", *J. Math. Phys.* **43**, 4537–4559 (2002).
[18]. J.S.Bell, "Bertlmann's socks and the nature of reality", *Journale de Physique* **42** 41–61 (1981); reprinted as Chapter 16 of *Speakable and Unspeakable in Quantum Mechanics*, Cambridge Univ. Press, (1987).
[19]. J.M.Keynes, *The general theory of employment, interest and money*, Macmillan, London, in preface, (1936).
[20]. G.T.Gillies, "The Newtonian gravitational constant: recent measurements and related studies", *Reports on Progress in Physics*, **60**, 151–225 (1997).
[21]. M.Fleischmann, S.Pons and M.Hawkins, "Electrochemically induced nuclear fusion of deuterium", *J. Electroanalytical Chemistry* **261**. 301–308 (1989).
[22]. E.Davenas,F.Beauvais, J.Arnara,M.Oberbaum, B.Robinzon,A.Miadonna, A.Tedeschi, B.Pomeranz, P.Fortner, P.Belon, J.Sainte-Laudy, B.Poitevin, J. Beneviste, "Human basophil degranulation triggerred by very dilute antiserum against IgE", *Nature* **333**, 816–818 (1988).
[23]. http://www.bbc.co.uk/info/report2004/pdfs/BBC_news_review.pdf, *British Broadcasting Corporation website*, (2005).
[24]. I.Asimov, *Foundation*, Weidenfeld and Nicolson, London (Vol. 1, part 3, chapter 7), (1953).

APPENDIX: THE COX DERIVATION OF PROBABILITY

Richard Cox's approach applies the logic of inference, not immediately to the arbitrary complexities of global problems, but merely to the simple, unambiguous, 8-state toy world of up to three binary switches. Remarkably, this tiny world suffices to define the rules of probability calculus. My brief outline here follows the original development of Cox [1,2], with commentary by Jaynes ([10], chapter 2). I have modernised the context, re-ordered some of the material with clarity in mind, and omitted the idiosyncratic manipulations required to solve the functional equations that occur. There are levels of pedantry beyond anything I attempt, but they don't change the conclusions.

Our belief about the state S of a system is always in a specific context X, and we write $\pi(S \mid X)$ for it. Thus, in the 1-bit context $I = \{\downarrow, \uparrow\}$ of a single switch A, we write our belief in A being "↑" as

$$(A \mid I) = \{\uparrow\} \mid \{\downarrow,\uparrow\}, \qquad \text{with belief } \pi(A \mid I)\,, \tag{1}$$

and the converse "not A" as

$$(\overline{A} \mid I) = \{\downarrow\} \mid \{\downarrow,\uparrow\}, \qquad \text{with belief } \pi(\overline{A} \mid I)\,. \tag{2}$$

In the 2-bit context $J = \{\downarrow\downarrow,\downarrow\uparrow,\uparrow\downarrow,\uparrow\uparrow\}$ of two switches A,B, we have

$$(A \mid J) = \{\uparrow\downarrow,\uparrow\uparrow\} \mid \{\downarrow\downarrow,\downarrow\uparrow,\uparrow\downarrow,\uparrow\uparrow\}, \ \text{with belief } \pi(A \mid J) \tag{3}$$

for the first bit A, and similarly for the second bit B. There are also "A AND B" joint beliefs about two bits both being "↑"

$$(AB \mid J) = \{\uparrow\uparrow\} \mid \{\downarrow\downarrow,\downarrow\uparrow,\uparrow\downarrow,\uparrow\uparrow\}, \qquad \text{with belief } \pi(AB \mid J)\,, \tag{4}$$

and other conditional beliefs such as

$$(B \mid AJ) = \{\uparrow\uparrow\} \mid \{\uparrow\downarrow,\uparrow\uparrow\}, \qquad \text{with belief } \pi(B \mid AJ)\,. \tag{5}$$

In the 3-bit context $K = \{\downarrow\downarrow\downarrow,\downarrow\downarrow\uparrow,\downarrow\uparrow\downarrow,\downarrow\uparrow\uparrow,\uparrow\downarrow\downarrow,\uparrow\downarrow\uparrow,\uparrow\uparrow\downarrow,\uparrow\uparrow\uparrow\}$ of A,B,C,

$$\begin{aligned}(A \mid K) \ &= \{\uparrow\downarrow\downarrow,\uparrow\downarrow\uparrow,\uparrow\uparrow\downarrow,\uparrow\uparrow\uparrow\} \mid \{\downarrow\downarrow\downarrow,\downarrow\downarrow\uparrow,\downarrow\uparrow\downarrow,\downarrow\uparrow\uparrow,\uparrow\downarrow\downarrow,\uparrow\downarrow\uparrow,\uparrow\uparrow\downarrow,\uparrow\uparrow\uparrow\} \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{with belief } \pi(A \mid K)\end{aligned} \tag{6}$$

and similarly for the other variants.

We aim to develop a calculus for manipulating our beliefs about this system, and start by **asserting transitivity** — if, in context K, we have more belief in A than B, and more in B than C, then we assert that we have more belief in C than in A. To do otherwise would lead us to argue in circles (or at least triangles). A consequence is that we can map π (whatever it was originally) onto real numbers, in which "more belief in" is represented by ">". The transitivity assertion is

$$\left.\begin{array}{l}\pi(A \mid K) > \pi(B \mid K) \\ \pi(B \mid K) > \pi(C \mid K)\end{array}\right\} \Longrightarrow \pi(A \mid K) > \pi(C \mid K) \tag{7}$$

So **beliefs are real numbers** — or at least they may as well be.

We now **assert that** knowing about A, and also about B given A, suffices to teach us about AB, all in the same overall context J. Some function F depending on both its arguments formalises this inference:

$$\begin{array}{ll}(A \mid J) = \{\uparrow\downarrow,\uparrow\uparrow\} \mid \{\downarrow\downarrow,\downarrow\uparrow,\uparrow\downarrow,\uparrow\uparrow\}, & \text{belief } a = \pi(A \mid J)\,, \\ (B \mid AJ) = \{\uparrow\uparrow\} \mid \{\uparrow\downarrow,\uparrow\uparrow\}, & \text{belief } b = \pi(B \mid AJ)\,, \\ (AB \mid J) = \{\uparrow\uparrow\} \mid \{\downarrow\downarrow,\downarrow\uparrow,\uparrow\downarrow,\uparrow\uparrow\}, & \text{belief } \pi(AB \mid J) = F(a,b)\,.\end{array} \tag{8}$$

The world of three bits allows sequential learning too. The three beliefs

$$\begin{array}{ll}(A \mid K) = \{\uparrow\downarrow\downarrow,\uparrow\downarrow\uparrow,\uparrow\uparrow\downarrow,\uparrow\uparrow\uparrow\} \mid \{\downarrow\downarrow\downarrow,\ldots,\uparrow\uparrow\uparrow\}, & \text{belief } x = \pi(A \mid K) \\ (B \mid AK) = \{\uparrow\uparrow\downarrow,\uparrow\uparrow\uparrow\} \mid \{\uparrow\downarrow\downarrow,\uparrow\downarrow\uparrow,\uparrow\uparrow\downarrow,\uparrow\uparrow\uparrow\}, & \text{belief } y = \pi(B \mid AK) \\ (C \mid ABK) = \{\uparrow\uparrow\uparrow\} \mid \{\uparrow\uparrow\downarrow,\uparrow\uparrow\uparrow\}, & \text{belief } z = \pi(C \mid ABK)\end{array} \tag{9}$$

chain together to define our belief $\pi(ABC \mid K)$. In the chain, B could be combined with A before linking with C, or with C before A. Hence

$$\pi(ABC \mid K) = F(\underset{AB|K}{F(x,y)}, \underset{C|ABK}{z}) = F(\underset{A|K}{x}, \underset{BC|AK}{F(y,z)}) \tag{10}$$

in which the arguments of the outer F's are labelled underneath. This is the "associativity equation" and it restricts F to be of the form

$$F(a,b) = G^{-1}\big(G(a) + G(b)\big) \tag{11}$$

where G is some invertable function of only one variable, instead of two. (This is often quoted multiplicatively as $W^{-1}(W(a)W(b))$, but it's better here not to exponentiate in the proof, and let the components add instead.) Remembering that π was initially on an arbitrary scale, we can upgrade to a less arbitrary scale of belief

$$\phi(\cdot) = G(\pi(\cdot)) \tag{12}$$

in which sequential learning (8) proceeds by addition

$$\phi(AB \mid J) = \phi(A \mid J) + \phi(B \mid AJ). \tag{13}$$

Yet there remains some arbitrariness, because ϕ could be rescaled by any function ρ to $\rho(\phi(\cdot))$ provided

$$\rho(u+v) = \rho(u) + \rho(v) \tag{14}$$

so that (13) still holds. Any constant scaling $\rho(\phi) = \gamma\phi$ suffices, although that is the only such freedom. To fix the scale completely, we consider negation.

In the two-state world $I = \{\downarrow, \uparrow\}$ of a single bit, we **assert that** our belief about A defines our belief about its converse $\overline{A}$, formalised by some function f

$$\phi(\overline{A} \mid I) = f\big(\phi(A \mid I)\big). \tag{15}$$

Repeated negation is the identity, so

$$f\big(f(x)\big) = x. \tag{16}$$

Now consider the three-state world $T = \{\downarrow\uparrow, \uparrow\downarrow, \uparrow\uparrow\}$ of two bits A and B in which at least one is "$\uparrow$". With context T understood throughout,

$$\begin{aligned}
\phi(AB) &= \phi(A) + \phi(B \mid A) && \text{sequential learning} \\
&= \phi(A) + f\big(\phi(\overline{B} \mid A)\big) && \text{definition of } f \\
&= \phi(A) + f\big(\phi(\overline{B}, A) - \phi(A)\big) && \text{sequential de-learning} \\
&= \phi(A) + f\big(\phi(\overline{B}) - \phi(A)\big) && B = \downarrow \text{ state is unique in } T \\
&= \phi(A) + f\big(f(\phi(B)) - \phi(A)\big) && \text{definition of } f \\
&= x + f\big(f(y) - x\big) && \text{name } \phi(A) = x,\ \phi(B) = y.
\end{aligned} \tag{17}$$

Symmetry $\phi(AB \mid T) = \phi(BA \mid T)$ then gives

$$x + f\big(f(y) - x\big) = y + f\big(f(x) - y\big) \tag{18}$$

The functional equations (16) and (18) together require

$$f(x) = \gamma^{-1} \log(1 - e^{\gamma x}) \tag{19}$$

and hence

$$\exp\left(\gamma\phi(\overline{A} \mid I)\right) = 1 - \exp\left(\gamma\phi(A \mid I)\right) \tag{20}$$

Remembering that ϕ could be scaled with any coefficient γ, we can upgrade to a fixed scale by defining $\Pr(\cdot) = \exp(\gamma\phi(\cdot))$. Qualitatively, $0 \leq \Pr(\cdot) \leq 1$ because neither exponential in (20) can be negative. Quantitatively, (20) becomes

$$\Pr(A \mid I) + \Pr(\overline{A} \mid I) = 1 \tag{21}$$

which is the sum rule. Meanwhile, (13) exponentiates to

$$\Pr(AB \mid J) = \Pr(A \mid J)\,\Pr(B \mid AJ) \tag{22}$$

which is the product rule. We have derived the **sum and product rules** of probability calculus, and there's no scaling freedom left. With both rules obeyed, we are entitled to call $\Pr(S \mid X)$ the **probability** of state S in context X.

For general inference, we use the simple switches $A, B, C, \ldots$ to encode arbitrary propositions. Applying the product rule when we know B to be "↑" (hence $\Pr(AB \mid J) = \Pr(A \mid J)$), shows that the true statement $(B \mid AJ)$ has $\Pr(B \mid AJ) = 1$. In general context, the unique true proposition thus has to be assigned $\Pr(\texttt{TRUE}) = 1$. The negation of truth being falsity, it follows from the sum rule that $\Pr(\texttt{FALSE}) = 0$. Hence

$$\Pr(\texttt{FALSE}) = 0 \leq \Pr(\cdot) \leq 1 = \Pr(\texttt{TRUE}) \tag{23}$$

and this restricts the assignments we can make.

The Cox derivation rests only upon elementary logic applied to very small worlds. If there is a general theory of rational inference at all, it must apply in special cases, so it can only be this probability calculus. Moreover, any defined problem can be broken down into small steps. We have to use probability calculus in the small steps, and this implies using it overall in the larger problem. This is the only globally-applicable calculus we are ever going to have, so we should use it. And it does seem to be rather successful.