I was a Teenage Logical Positivist (Now a Septuagenarian Radical Probabilist)

Dick Jeffrey's PSA Presidential Address
Kansas City, October 23rd 1998
Last tweaked on...

April 10, 2000

http://www.princeton.edu/~bayesway/KC.tex.pdf

In memory of Peter (C.G.) Hempel

January 8th 1905–November 7th 1997

1 Contents

Mostly, modifications of transparencies used in the talk:

2 The kid reads Carnap

—in Boston, 1943, in E.S. Brightman's metaphysics seminar:

> *Philosophy and Logical Syntax,*
> 100 pp., Psyche Miniatures, 1935

"These are my people!"

[Brightman was a "Personalist" (a.k.a. "Personal Idealist"), i.e., he held that only persons are truly real. He mentioned Carnap by the way as the leading logical positivist.]

—and Chicago, 1948, in Carnap's seminar:

> *The Logical Syntax of Language,*
> International Library of Philosophy
> Psychology and Scientific Method, 1937

[In Chicago right after the war the kid was amazed to find the philosophy department dominated by a certain Richard McKeon (cf. Pirsig 1974), whom he had never heard of, and whose acolytes regarded Carnap as a bad joke. But others of us recognized Carnap as a treasure: Ruth Marcus (then a post-doc), a number of U of C students (Bill Alston, Norman Martin, Howard Stein, Bob Palter, Stan Tennenbaum, ...), and my guru Abner Shimony, who was then a young visiting student from Yale, deeply into Whitehead's metaphysics.

Reference: Robert M. Pirsig, *Zen and the Art of Motorcycle Maintenance* (1974)

3 Logical Empiricism, "Scientific Philosophy"

It was not a position but a movement—with activists in Vienna, Berlin, Prague, Warsaw, Uppsala, London, ..., and even Cambridge (Massachusetts).

The shape of the movement was significantly influenced by historical contingencies like the execution by the Gestapo in Vilna of Janina Hosiasson-Lindenbaum (1899-1942), silencing one of probabilism's clearest early voices.

Some of her writings:

"Why do we prefer probabilities relative to many data?", Mind **40** (1931) 23-36

"On confirmation", *Journal of Symbolic Logic* (1940) 133-148

"Induction et analogie: comparaison de leur fondement", *Mind* (1941) 351-365

The movement was sparked by Russell, *Our Knowledge of the External World as a Field for Scientific Method in Philosophy* (1915). ("Scientific Method in Philosophy" was the running head and title on the cover of the first edition.) *Der logische Aufbau der Welt* (1928) was Carnap's attempt to implement Russell's program.

4 Logical Empiricism as a Position

In “Two dogmas of empiricism” (with “logical” understood) Quine took empiricism to be Carnap’s (1928) phenomenalistic reductionism—coupled with the despised analytic/synthetic distinction.

> [Carnap had long since abandoned the 1928 idea in favor of the “radical physicalism” floated by Neurath and himself in the protocol-sentence debate (*Erkenntnis,* ca. 1932), and by him in part V of *Logische Syntax der Sprache* (1934).]

The *neat* thing was the logicism that made sense of the non-empirical character of mathematics by rooting it in logic.

What happened to logical empiricism, the position?

Did logicism crash?

  No, it just got leaner and cleaner (**5, 6, 7**).

Did empiricism crash?

  Yes, i.e., empiricist epistemology (**8**).

Rudolf Carnap, *Der logische Aufbau der Welt* (1928), translated as *The Logical Structure of the World* (1967)

5 Proto-logicism: Numbering Finite Sets

Q: How do the natural numbers latch onto the world?

A: They don't; e.g., when we explicate

> There are (exactly) 3 dodos

we find dodos mentioned, but not numbers:

> There exist distinct dodos, x, y, z,
> and no others

—or, in logical notation,

$$\exists x\, \exists y\, \exists z [Dx \,\&\, Dy \,\&\, Dz \,\&\, x \neq y \,\&\, x \neq z \,\&\, y \neq z \\ \&\, \sim \exists w (Dw \,\&\, w \neq x \,\&\, w \neq y \,\&\, w \neq z)]$$

For short: $(\exists 3x) Dx$.

Aside from the biological term *dodo* (D), only logical terms appear here.

Reference: In the factual claim "$(\exists 3x)Dx$", the term "3" is an incomplete symbol (*Whitehead and Russell, Principia Mathematica* I, p. 66), defined away in the familiar contextual definition of the numerically definite quantifier illustrated above.

6 Logicism Lite: It's the *Data* that are Logical

Carnap's full-blown (1931) logicist thesis about what he called simply "mathematics":

1. Its concepts are explicitly definable in logical terms.
2. Its theorems then become logical truths.

A lighter logicism would count number-theoretical data as logical for the same reason that physical data are counted as empirical—i.e., because of how they are grounded. Theorems like Fermat's Last Theorem would count not as shorthand for logically valid formulas, but as generalizations that derive their status as logical from the data they are responsible to, e.g., data like invalidity of the schema

$$(\exists\, 2^3 + 3^3\, x)\, \phi x \equiv (\exists\, 5^3\, x)\, \phi x,$$

or, in a theoretically simpler notation,

$$(\exists\, 0''^{0'''} + 0'''^{0'''}\, x)\, \phi x \equiv (\exists\, 0'''''^{0'''}\, x)\, \phi x.$$

Such data are established via recursive contextual definitions of $+$, $\times$, etc. in numerically definite quantifiers—e.g., in the case of exponentiation, this pair:

$$\begin{aligned} &(\exists\, n^0\, x)\, \phi x \equiv (\exists\, 0'\, x)\, \phi x, \\ &(\exists\, n^{m'}\, x)\, \phi x \equiv (\exists\, n^m \times n\, x)\, \phi x \end{aligned}$$

Now any decision procedure for monadic 1st-order logic with '=' delivers the data about particular sums, products, and powers to which number theory is answerable.

7 Number Theory: Any ω-sequence Goes

If you want to do number theory, you need numbers to quantify over and to serve as referents of the numerals '0', '1', etc. *Definition:* An ω-sequence is a pair $\langle domain, function\rangle$ that satisfies the following ("Peano") axioms—in which the function is called "successor of".

1. All members of the domain have unique successors.

2. No member is the successor of more than 1 member.

3. 0 is not the successor of any member.

4. The principle of mathematical induction holds.

Comrades, any ω-sequence you like can play the role of the pair $\langle Natural\ numbers,\ successor\ of\rangle$. A likely candidate is the sequence of numerals, '0', '$0'$', '$0''$', . . ., paired with the function *append an accent.* (The syntax itself requires the existence of some such sequence.) Here the natural numbers are the numerals themselves, harmlessly self-referential.

But any ω-sequence will do. (And your choice need not be mine. Think: inverted spectra. And here, substitutional quantification does the job.) Here, choice is otiose.

Reference: R.C. Jeffrey, "Logicism 2000: A Mini-manifesto", *Benacerraf and his Critics* (1996) 160-164, Adam Morton and Stephen P. Stich (eds.)

8 It's Empiricism that Crashed

The sorts of things I have in mind:

• *Mach's moustache* frames an inert, physiologically ignorant photographic parable of visual perception:

> "[. . .]I lie upon my sofa. If I close my right eye, the picture represented in [Fig. 1] is presented to my left eye. In a frame formed by the ridge of my eyebrow, by my nose, and by my moustache, appears a part of my body, so far as visible, with its environment. *My* body differs from other human bodies [. . .] by the circumstance that it is only partly seen, and, especially, is seen without a head. [. . .]"
> —*Analysis of Sensations* (1890) 15-16

• *Quine's irritations* of our sensory surfaces—say, the retina. A chronological record of the pattern of irritation of rods and cones would be gibberish without a correlated record of activity in the proprioceptive circuitry monitoring position of eyes in head, head on torso, etc.

• *Conscious experience* is too slender a base, e.g., we are usefully sensitive to pheromones that we cannot smell.

• *Inner-outer hocus-pocus:* use of the familiar inside/outside the skin contrast as an explanatory placebo. The whole body is part of the "external" world.

• *Epistemology naturalized?* No. Probabilistic methodology is the way to go.

9 "Dogmatic" Probabilism

("Dogmatic": probabilities are grounded in certainties.)

C. I. Lewis's *Analysis of Knowledge and Valuation* (1946) was addressed in Carnap's seminar in 1948. On p. 186 Lewis enunciates the dogmatic probabilism to which I would oppose radical probabilism (probabilities all the way down to the roots; see the probability kinematics of **14** below):

> "If anything is to be probable, then something must be certain. The data which themselves support a genine probability, must themselves be certainties. We do have such absolute certainties in the sense data initiating belief and in those passages of experience which later may confirm it."

See also his *Mind and the World Order* (1929), pp. 328-9:

> "the immediate premises are, very likely, themselves only probable, and perhaps in turn based upon premises only probable. Unless this backward-leading chain comes to rest finally in certainty, no probability-judgment can be valid at all. [...] Such ultimate premises [...] must be actual given data for the individual who makes the judgment,"

A useful 3-way discussion of Lewis's idea appeared in *Philosophical Review* **61** (1952) 147-175.

10 Neurath on Dogmatic Protocols

A halfway house between epistemology and methodology:

"Protokollsaetze", *Erkenntnis* **3** (1932-3) 204-214

Translation in Otto Neurath, *Philosophical Papers 1913-1946*, Reidel (1983):

> "*There is no way to establish fully secured, neat protocol sentences as starting point of the sciences.* There is no *tabula rasa.* We are like sailors [...]" (P. 92)

> "[...] Fundamentally it makes no difference at all whether Kalon works with Kalon's or with Neurath's protocols [...]. In order to make this quite clear, we could think of a scientific cleaning machine into which protocol sentences are thrown. The 'laws' and other 'factual statements', including protocol statements, which have their effect through the arrangement of the wheels of the machine, clean the stock of protocol statements thrown in and make a bell ring when a 'contradiction' appears. Now either the protocol statement has to be replaced by another or the machine has to be rconstructed. *Who* reconstructs the machine, *whose* protocol statements are thrown in, is of no consequence at all; everybody can test his 'own' as well as 'others" protocol statements. (P. 98)

11 Moore-Shannon Protocol Enhancement

Driving to work, radios tuned to NPR, Ann and three of her colleagues all hear an actor—they know it's Gielgud or Olivier—doing "To be or not to be". On arrival they write protocol sentences on cards (e.g., "Ann's protocol at 9 AM: At 8:45 AM I heard Gielgud") and drop them into the protocol box. The *Protokollmeister* collects the 4 cards & prepares a single protocol for the Neurath machine (e.g., "Master protocol at 9:05 AM: "It was Gielguid"), like this:

> The master protocol says it was Gielgud if at least 3 of the individual protocols said it was Gielgud, and otherwise says it was Olivier.

The *Protokollmeister* regards the four individuals as equally reliable—and as not very reliable. He thinks they are all pretty good at recognizing Gielgud's voice, and really bad at recognizing Olivier's. For each of them—say, Ann—he judges:

$pr(\text{Ann says "Gielgud"}|\text{It is Gielgud}) = 80\%$
$pr(\text{Ann says "Gielgud"}|\text{It is Olivier}) = 60\%$

Fact: He must judge that the master protocol, *MP,* does better:

$pr(MP \text{ says "Gielgud" } | \text{ It is Gielgud}){=}82\%$
$pr(MP \text{ says "Gielgud" } | \text{ It is Olivier}){=}47\%$

Proof. The probability of exactly 3 "Gielgud"s is $4p^3(1-p)$, where $p = 80\%$ if it was Gielgud, and $p = 60\%$ if not. Then the probability of 3 or 4 "Gielgud"s is $4p^3(1-p)+p^4$, which $= p^3[4(1-p)+p]$.

If $p = 80\%$, this $= (.512)[(4)(.2) + .8] = 82\%$.
If $p = 60\%$, this $= (.216)[(4)(.4) + .6] = 47\%$.

And many more 80%/60% protocols can make a 99%/1% one.

Reference: Edward Moore & Claude Shannon, "Reliable Circuits Using Less Reliable Relays", *J. Fanklin Inst.* **262** (1956) 191-297

12 NOT "the definition" of P(B|A):

$$P(B|A) = \frac{P(AB)}{P(A)} \text{ when } P(A) > 0$$

There are two reasons for this:

- $P(B|A)$ may be defined when $P(A) = 0$ in idealized examples:
  $P(\text{The } H_2O \text{ is solid} \mid \text{Its temperature is precisely } \pi^\circ F) = 1$

- $P(B|A)$ may be defined when $P(AB)$ and $P(A)$ are not.

Q: What *is* the definition, then?

A: There is none, any more than there is an impersonal definition, "the" definition, determining our unconditional probabilities.

Conditional and unconditional probability are different functions, of different numbers of arguments: 2 and 1. They are connected by the multiplicative law,

$$P(AB) = P(A)P(B|A),$$

which can be solved for $P(B|A)$ only in cases where $P(A)$ and $P(AB)$ are defined and positive.

13 When is Conditioning the Way to Go?

Notation. P, Q: your probability functions $before$, $after$ updating

Question: When is it OK for you to set $Q(H) = P(H|data)$?

Necessary Condition #1, "Certainty": $Q(data) = 1$

Certainty is not sufficient—e.g., because when it's OK to condition on $data$ it may not be OK to condition on logical consequences $data'$ of $data$, even though $Q(data') = Q(data) = 1$.

EXAMPLE

$data$: 'The card is a heart'

$data'$: 'The card is red'

P(It's the queen of hearts | It's a heart)$= \frac{1}{13}$, but

P(It's the queen of hearts | It's red)$= \frac{1}{26}$

Necessary Condition #2, in three equivalent versions:

- *Constant Odds:*

  If B and C each imply $data$, then $\dfrac{Q(B)}{Q(C)} = \dfrac{P(B)}{P(C)}$.

- *Constant Proportion:*

  If B implies $data$, then $\dfrac{Q(B)}{Q(data)} = \dfrac{P(B)}{P(data)}$.

- *Rigidity:*

  $Q(H|data) = P(H|data)$ for all H.

One way to ensure that these three hold: Use a "statistician's stooge" (I. J. Good's term), i.e., someone you trust to give true yes-or-no answers, with no further information, to questions of form "$data$?".

14 Probability: Protocols and Kinematics

NOTATION & JARGON

- $D_1, \ldots D_n$: The best possible outcomes of a certain observation, forming a "partition" (exhaustive and mutually exclusive).
- P: your old probability function, to be modified by the observation.
- Q: your new probability function, as modified by the observation.
- *Conditioning* in response to the protocol D_i:

  $Q(H) = P(H|D_i)$ ("Conditioning")
- *Probability Kinematics* (a.k.a. "Jeffrey conditioning") in response to probability protocols $Q(D_i) = q_i$:

(1) $$Q(H) = \sum_{i=1}^{n} q_i P(H|D_i) \quad \text{("Probability Kinematics")}$$

Formula (1) (Jeffrey 1957) was a Carnap-style *proposal* about updating in cases where conditioning would work if only some q_i were 1.

Q: How are those cases to be identified? (Unasked in 1957)

A: By *Rigidity* (Jeffrey 1965) of $D_1, \ldots, D_n$ relative to $\{P, Q\}$ (a.k.a. *Sufficiency*: D&Z 1982):

(2) $Q(H|D_i) = P(H|D_i), = c_i$ for short ("Rigidity")

References. R.C. Jeffrey: *Contributions to the Theory of Inductive Probability* (Ph. D. thesis, Princeton, 1957); *The Logic of Decision* (1965, 1983, 1990), ch. 11; *Probability and the Art of Judgment* (1992), ch. 1, 6, 7. "D&Z", Persi Diaconis & Sandy Zabell: "Updating Subjective Probability", J. Am. Stat. Assn. **77** (1982) 822-830.

15 Solo Updating

Equations (1)-(5) are equivalent. Note that conditions (2)-(4) are recycled from the case (**13**) where certainty holds.

(1) $Q(H) = \sum_{i=1}^{n} q_i c_i,$ *Probability Kinematics*
where $q_i =_{df} Q(D_i)$ and $c_i =_{df} P(H|D_i)$

(2) $Q(H|D_i) = c_i$ for $i = 1, \ldots, n.$ *Rigidity*

(3) $\dfrac{Q(B)}{Q(C)} = \dfrac{P(B)}{P(C)}$ if $B, C \subset D_i$ *Constant Odds*

(4) $\dfrac{Q(B)}{Q(D_i)} = \dfrac{P(B)}{P(D_i)}$ if $B \subset D_i$ *Constant Proportion*

Definition, relative to the update $P \mapsto Q$:

$$f(H, G) =_{df} \frac{Q(H)}{P(H)} / \frac{Q(G)}{P(G)}$$ *Bayes factor*

—i.e., what you can multiply your old odds by to get your new odds. For a partition element $D_i \in \{D_1 \ldots, D_n\}$, we abbreviate:

$$f_i =_{df} f(D_i, D_n) = \frac{q_i}{p_i} / \frac{q_n}{p_n}$$ *Factor protocol*

—where the choice of D_n (instead of D_1, say) is arbitrary. Formula (1) can now be rewritten as follows:

(5) $Q(H) = \dfrac{\sum_{i=1}^{n} c_i p_i f_i}{\sum_{i=1}^{n} p_i f_i}$ *Factor Kinematics*

—which would have been no different if we had set $f_i =_{df} f(D_i, D_j)$ for any other fixed partition element D_j. Finally, note that with $H = D_i$ in (5) we have

(6) $q_i = \dfrac{p_i f_i}{\sum_{i=1}^{n} p_i f_i}$

16 Updating on an Expert's Factor Protocols

Factor protocols f_i can be better candidates than probability protocols q_i for the role of reports to be pooled, Neurath-style, in the common data base, for the f_i's seem to dissect out of the transition $P \mapsto Q$ the contribution of the observation itself, leaving the observer's prior probabilities behind.

MEDICAL EXAMPLE. In the light of a histopathologist's factor protocols, you, the clinician, update your prior probability $P(H)$ for your patient's being alive in 5 years. The partition has three cells:

(D_1) Islet cell ca, (D_2) Ductal cell ca, (D_3) Benign tumor.

Pathologist's (***bold***) probabilities $\boldsymbol{p_i, q_i}$ and factors $\boldsymbol{f_i}$ $(i = 1, 2, 3)$:

Probabilities before biopsy: $\boldsymbol{p_i} = P_{path}(D_i) = \frac{1}{2}, \frac{1}{4}, \frac{1}{4}$

Probabilities after biopsy: $\boldsymbol{q_i} = Q_{path}(D_i) = \frac{1}{3}, \frac{1}{6}, \frac{1}{2}$

Bayes factors: $\boldsymbol{f_i = (q_i/p_i)/(q_3/p_3)} = \frac{1}{2} \times \boldsymbol{q_i/p_i} = \frac{1}{3}, \frac{1}{3}, 1$

Clinician's (your) probabilities, $p_i, q_i, c_i, P_{clin}(H), Q_{clin}(H)$:

Before biopsy: $p_i = P_{clin}(D_i) = \frac{1}{6}, \frac{1}{6}, \frac{2}{3}$

After biopsy: $q_i = \frac{p_i f_i}{\sum p_i f_i} = Q_{clin}(D_i) = \frac{1}{14}, \frac{1}{14}, \frac{12}{14}$

Before = after: $c_i = P_{clin}(H|D_i) = Q_{clin}(H|D_i) = .4, .6, .9$

Before: $P_{clin}(H) = \sum_{i=1}^{3} c_i p_i \approx 77\%$ (Law of Total Probability)

After: $Q_{clin}(H) = \frac{1}{14/18} \sum_{i=1}^{3} c_i p_i f_i \approx 61\%$ (Factor kinematics)

Reference: Schwartz, Wolfe, and Pauker, "Pathology and Probabilities: a new approach to interpreting and reporting biopsies", *New England Journal of Medicine* **305** (1981) 917-923. (*Not* their numbers.)

17 Two Experts, One Partition

• In factor updating twice on the same partition, order is irrelevant:

MEDICAL EXAMPLE, CONTINUED. Adopting the pathologist's factor protocols $(f_i; i = 1, \ldots, n)$ as your own, you have updated your probabilities $P(D_i) = p_i$ on the diagnoses to $Q(D_i) = q_i$ as follows:

$$P \overset{f}{\mapsto} Q, \text{ where } q_i = \frac{p_i f_i}{\sum_{i=1}^n p_i f_i}$$

You also have factor protocols $(g_i; i = 1, \ldots, n)$ on the same partition of diagnoses from another sort of expert—say, a radiologist. Adopting those, you now update the q_i's to new probabilities $S(D_i) = s_i$:

$$P \overset{f}{\mapsto} Q \overset{g}{\mapsto} S, \ s_i = \frac{q_i g_i}{\sum_{i=1}^n q_i g_i} = \frac{\frac{p_i f_i}{\sum p_i f_i} g_i}{\sum(\frac{p_i f_i}{\sum p_i f_i} g_i)} = \frac{p_i f_i g_i}{\sum_{i=1}^n p_i f_i g_i}$$

Perhaps the two experts reported their protocols at the same time, but, inevitably, you have adopted them in a particular order, i.e., $P \overset{f}{\mapsto} Q \overset{g}{\mapsto} S$. But if you had adopted them in the other order, first using the radiologist's g_i's to update the $P(D_i) = p_i$'s to new values $R(D_i) = r_i = p_i g_i / \sum_{i=1}^n p_i g_i$ and only then using the pathologist's f_i's to update the r_i's to new values $T(D_i) = t_i$, the result would have been the same:

$$P \overset{g}{\mapsto} R \overset{f}{\mapsto} T, \ t_i = \frac{r_i f_i}{\sum_{i=1}^n r_i f_i} = \frac{\frac{p_i g_i}{\sum p_i g_i} f_i}{\sum(\frac{p_i g_i}{\sum p_i g_i} f_i)} = \frac{p_i g_i f_i}{\sum_{i=1}^n p_i g_i f_i} = s_i$$

And it would have been the same if you had updated in a single step, using the product protocols $f_i g_i$:

$$P \overset{fg}{\longmapsto} U, \text{ where } u_i = \frac{p_i f_i g_i}{\sum_{i=1}^n p_i f_i g_i} = t_i = s_i$$

18 Neurath's Machine Revisited

Neurath's machine (**10**) was envisioned in (**11**) as a box into which observers drop their protocols, whereupon a *Protokollmeister* filters its contents, replacing certain groups of unreliable protocols by more reliable single master protocols. Neurath himself spoke of the box as a machine with a *"Contradiction!"* bell that rings when the data base must be revised—but by no anointed *Protokollmeiser*.

Neurath's protocols were imperfectly reliable flat statements, in need of screening. But the probabilistic protocols of **16** and **17** may also need screening and revision (Garber 1980).

EXAMPLE: DOUBLE BILLING? When you adopt the pathologist's factor protocols "$\frac{1}{3}, \frac{1}{3}, 1$" in **16** you update your 2:1 odds on *benign tumor* (derived from your 1:1:4 odds on D_1, D_2, D_3) to 6:1. Now suppose you ask for a second opinion, and get "$\frac{1}{3}, \frac{1}{3}, 1$" again. A simple soul would update 6:1 further, to 18:1. But

(a) if you take the second three factors to represent the independent opinion of a second pathologist you will not update beyond 6:1, for you take the second opinion to merely endorse the first; and

(b) if you see the "second" opinion as just bogus reiteration by the first pathologist, you may mistrust the first report, and return to your original 2:1.

Two morals:

• It is always your own protocols that you update on. ("Montaigne: "Even on the world's highest throne we sit on our own bottom.") You need not adopt expert protocols. You may modify them:

$$\text{(a)}\ P \overset{f}{\longmapsto} Q \overset{1}{\longmapsto} Q \qquad\qquad \text{(b)}\ P \overset{f}{\longmapsto} Q \overset{f^{-1}}{\longmapsto} P$$

• Consultation is sounder than faxed opinion as a basis for judgment.

References: Daniel Garber, "Field and Jeffrey Conditionalization", *Philosophy of Science* **47** (1980) 142-145. Montaigne, *Essays* (1588), next-to-last paragraph (i.e., in the essay "On Experience").

19 Two Experts, Two Partitions

Factor updating need not be commutative when factors are for different partitions; and here probability updating *can* be commutative.

Updating with new probability protocols $a = \langle a_1, \ldots, a_n \rangle$ on the partition $\langle A_i, \ldots, A_n \rangle$, and $b = \langle b_1, \ldots, b_n \rangle$ on the partition $\langle B_i, \ldots, B_m \rangle$, D&Z define P_a, P_{ab}, etc. as follows:

$$P \overset{a}{\longmapsto} P_a,\ P \overset{b}{\longmapsto} P_b,\ P_a \overset{b}{\longmapsto} P_{ab},\ P_b \overset{a}{\longmapsto} P_{ba}$$

They show that commutativity is tied to *independence,* in two senses: They call the A and B partitions...

- "$P-$independent" iff for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$, $P(A_i|B_j) = P(A_i)$ and $P(B_j|A_i) = P(B_j)$, and
- "$J-$independent" relative to P, a, b iff for all such i, j, $P_a(B_j) = P(B_j)$ and $P_b(A_i) = P(A_i)$

D&Z prove that $P_{ab} = P_{ba} \cdots$

(3.1) if for all A_i, B_j, $P_{ab}(A_i) = P_a(A_i)$, $P_{ba}(B_j) = P_b(B_j)$, and

(3.2) iff the A and B partitions are $J-$independent relative to P, a, b.

If one of the partitions A, B has only two elements, $J-$independence relative to any particular pair a, b is equivalent to $P-$independence; but in general,

(3.3) $P-$independence is equivalent to J-independence relative to P, a, b for *all* a, b.

If you are to have the pre-assigned probabilities a, b on A, B after updating twice, the protocols must commute ($P_{ab} = P_{ba}$).

These facts about probability protocols apply indirectly to factor protocols, since the former are derivable from the latter together with with the priors. (See also sec. 4 of D&Z.)

Reference: D&Z, Diaconis and Zabell, *JASA* (1982), esp. pp. 825-827.

20 On Medical Experience

Probability protocols may be the way to go when the A, B partitions are thought of as the sorts of things that the (differently) trained eyes of the two sorts experts might be sensitive to, i.e., not sense data but physical states that other experts would also be sensitive to and assign probability protocols to in the light of the observation. These A_i's and B_j's may be treated as hidden epistemological variables, which we may be convinced are independent even though the experts (say, two pathologists, looking through the eyepieces of the same microscope) cannot give them clear linguistic expression outside the immediate context of shared observation.

This is medical experience in Galen's sense: hands on, walking the wards, where the role of sense data is played by indications that anyone might see or smell or feel, and might prompt an expert (even if less than certain of the relevant character of the indications) to adopt protocols like $P_a(benign) = \frac{1}{2}$ as in **16**.

If such protocols are seen as unconnected to prior judgments (and indeed the protocol is usable even if the expert's P function was undefined for the value *benign*), then probability protocols may adequately represent what is learned from experience. In this case the background of the judgments at the end of **18** might be discussed in terms of dependency between the experts' experiences; and cases where commutativity holds might be understood in terms of judgments of independence because of (say) physical separation—judgments requiring no familiarity with the particular A_i's.

(Needs fleshing out.)

Reference. Galen, *Three Treatises on the Nature of Science*, translated by Richard Walzer and Michael Frede, with an introduction by Michael Frede (1984)

21 Mr Natural Explains it All

Kid: What does it MEAN, Mr Natural, when $P_{ab} \neq P_{ba}$?

Mr N: It means that the observations are not independent; one of them gives you some of the information that the other does.

Kid: Gosh, Mr Natural, does that solve Garber's problem about how enough repetitions of the same almost uninformative observation like $f(D_1) = 1.001$ can add up to a real informative one with a Bayes factor like 10?

Mr N: You got it, Kid. If you're sure the first has given you all the information you'll get from the rest, then $f(D_1) = 1.001$ only for the first; after that, the factors are all 1.

Kid: Gee willikers, Mr Natural, this is real exciting. If you've got a couple of swell observers who are confident of themselves and each other, does that mean it's like Garbersville?

Mr N: No, Kid. If they are totally confident then it's Certainty City, not Garbersville. And if they do have any doubt they can still learn from each other if initially they see their judgments as conditionally independent given the truth.

Kid: Oh wowsie, I'm totally excited. Does it all mean that non-commutativity is like NORMAL?

Mr N: That's an affirmative, Kid. And it means that normally, unless independence has been designed into the observations, factor protocols on a single partition will be inappropriate since they always commute.

Kid: [Wets pants]

22 Beyond Generalized Conditioning

Where the rigidity conditions are not satisfied by your upcoming observation, generalized conditioning is no way to go. But there may be other moves you can make:

(1) *The Problem of New Explanation* (Glymour: *Old Evidence*). Here, seemingly, $P(H|evidence) = P(H)$. Example: Einstein's discovery that the already well known anomalous advance in the perihelion of Mercury was explicable by the GTR. A sometimes applicable solution is "probability reparation" (Jeffrey; see Wagner for further progress and references.)

(2) *Temporal Coherence, Reflection; "Condition M"* (Goldstein, van Fraassen; Skyrms): "Current probability judgments are current expectations of future probability judgments." This condition holds wherever rigidity holds for an observable partition, and sometimes where it holds for none.

(3) *Beyond Condition M: Expected Irrationality.* Arriving at the party, you give your car keys to a friend to hold until you are home—because you anticipate that later in the evening it will be your drunken wish to drive home. This wise choice violates temporal coherence and reflection. (So does probability reparation.)

References

Clark Glymour: *Theory and Evidence*, 1980.

Michael Goldstein: *The Prevision of a Prevision JASA* **78** (1983) 231-248. "Prior Inferences for Posterior Judgments", *Structures and Norms in Science*, ed. Maria Luisa dalla Chiara *et al.* (1997) 55-72.

Richard Jeffrey, *Probability & the Art of Judgment* (1992) 103-107

Brian Skyrms, *The Dynamics of Rational Deliberation* (1990)

Bas van Fraassen: "Belief and the Will", *J. Phil.* **81** (1984) 235-256.

Carl Wagner: "Old Evidence and New Explanation", *Philosophy of Science* **64** (1997) 677-691.

23 Gappy Probabilities; Fine Points for Mavens

It's OK to have gappy probability assignments—e.g., because...

(1) In updating $P \mapsto Q$ by probability kinematics, you make no use of the prior $P(D_i)$'s, which can therefore be undefined.

(2) In evaluating your desirability for an act, i.e., your conditional expectation of utility given the act, you make no use of the prior probability of the act, which is best left undefined. [Then the prior probabilities of states of nature must also be left undefined, for, as Spohn observes,

$$P(A) = \frac{P(B) - P(B|\bar{A})}{P(B|A) - P(B|\bar{A})} \text{ if } P(B|A) \neq P(B|\bar{A}),$$

so that if state B is not P-independent of act A, and $P(B|A)$ and $P(B|\bar{A})$ are both defined, then $P(B)$ is undefined if $P(A)$ is.

(3) Your judgment $P(AB) = P(A)P(B)$ of independence can make sense even when you have no P values in mind.

(4) At an early stage of deliberation, old and new probabilities P, Q may be "entangled" in the sense that although you have not yet set numerical values for the $P(H|D_i)$'s, you do take the rigidity conditions $Q(H|D_i) = P(H|D_i)$ to hold. Setting values $P(H|D_i) = a_i$ disentangles Q and P, for you then have separate conditions, $P(H|D_i) = a_i$ on P and $Q(H|D_i) = a_i$ on Q, which jointly imply rigidity.

References

Wolfgang Spohn, *Erkenntnis* **11** (1977) 114-116.

Isaac Levi, *The Covenant of Reason* (1997) 73-83.

24 Mad-Dog Subjectivism

(1) There are no "real" probabilities out there for us to track,

(2) nor are there uniform probability-makers (cf. truth-makers),

(3) but there can be averages or symmetries out there, in view of which certain judgmental probability assignments are irresistible,

(4) as happens when we use probabilistic theories—notably, quantum mechanics. The probabilities it provides are "subjective" for us, i.e., we adopt them as our judgmental probabilities, but they are objective in the sense (3) of being shared and compelling.

> Kid: Is it all in the mind, Mr Natural?
>
> Mr N: All is cool, kid. We don't place the spots on the photographic plate. It's physics that's a social construct, not the physical world.

—That's All For Now, Folks—

This is pure de Finetti, going back to his "Probabilismo" (1931). For more about it, with references (to Haim Gaifman and others), see Bas van Fraassen, *Laws and Symmetries* (1989), e.g., pp. 198-199 regarding (4), and elsewhere in the vicinity for (1)-(3).