Experimental Results I: The Questionnaires

- There are many versions (dating back to K&T's [18]) of the questionnaires underlying the controversies about base rates. Here are Gigerenzer's [11]:

"Probability Format"	"Frequency Format"
The probability of breast cancer is 1% for a woman at age forty who participates in routine screening.	**10 out of every 1,000 women at age forty who participates in routine screening have breast cancer.**
If a woman has breast cancer, the probability is 80% that she will get a positive mammogram.[a]	8 out of every 10 women with breast cancer will get a positive mammogram.
If a woman doesn't have breast cancer, the probability is 9.6% that she will get a positive mammogram.	95 out of every 990 women without breast cancer will also get a positive mammogram.
A woman in this age group had a positive mammogram in a routine screening. What is the probability that she actually has breast cancer? ________ %	Here is a new representative sample of women at age forty who got a positive mammogram in routine screening. How many of these women do you expect to actually have breast cancer? _____ of _____

[a] Presumably, the subjects are "supposed to" interpret *conditionals* of the form ⌜If p, then $\Pr(q) = x$⌝ as *conditional probability statements* of the form ⌜$\Pr(q \mid p) = x$⌝. But, this is a *bad idea*. Consider $p = q = A$.

Experimental Results II: Patterns of Response

- Typically, subjects in these kinds of experiments are given a questionnaire in "probability format" (or in an analogous relative frequency format, which seems to elicit similar patterns of responses [11]).
- There have been *many* similar problems discussed over the years: "lawyer-engineer" problems [18], "taxicab" problems [30], [1], "light-bulb" problems [23], and disease problems [7], [5]. See [11] for a survey.
- In the (diagnostic) problems at hand, there is an interesting pattern observed in the responses given by subjects on the "probability format" questionnaire: answers (even of experts) tend to be between 70% – 80%.
- Moreover, patterns of response for the "frequency format" questionnaire tend to be significantly different. Experiments of Gigerenzer *et al* indicate a significantly lower average (*i.e.*, significantly < 70%) response [11].
- Which pattern is more "consistent" with "the probability calculus"?

What Does "The Probability Calculus" Say? Part I

- It is typically claimed (see, *e.g.*, [11]) that the "correct" probabilistic answer to this question (in "probability format") can be computed as follows.
- Let E = the evidence that a woman w (where w is *in the salient reference class*: $R = w \in \mathcal{R}$) has received a positive mammogram, and H = the hypothesis that w has breast cancer. Then, Bayes' theorem "says":

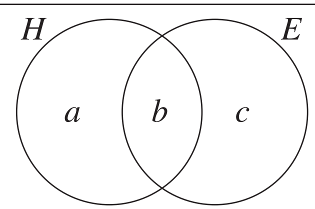
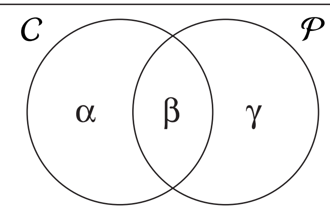
$$(1)\quad \Pr_R(H \mid E) = \frac{\Pr_R(E \mid H) \cdot \Pr_R(H)}{\Pr_R(E \mid H) \cdot \Pr_R(H) + \Pr_R(E \mid \neg H) \cdot \Pr_R(\neg H)} = \frac{(.80) \cdot (.01)}{(.80) \cdot (.01) + (.096) \cdot (.99)} = 0.078$$

- In "frequency format," one *can* (but *should* one? [26, ch. 1]) estimate the frequency of women in the target population ($\mathcal{T}$) who will have cancer ($\mathcal{C}$), among those having a positive test result ($\mathcal{P}$), as the corresponding frequency in the sample population ($\mathcal{S}$) [here, $\#_{\mathcal{Y}}(\mathcal{X})$ = # of elements in $\mathcal{X} \cap \mathcal{Y}$]:

$$(2)\quad f_{\mathcal{T}}(\mathcal{C} \mid \mathcal{P}) \doteq f_{\mathcal{S}}(\mathcal{C} \mid \mathcal{P}) = \frac{\#_{\mathcal{S}}(\mathcal{C} \cap \mathcal{P})}{\#_{\mathcal{S}}(\mathcal{P})} = \frac{\#_{\mathcal{S}}(\mathcal{C} \cap \mathcal{P})}{\#_{\mathcal{S}}(\mathcal{P} \cap \mathcal{C}) + \#_{\mathcal{S}}(\mathcal{P} \cap \overline{\mathcal{C}})} = \frac{8}{8 + 95} = 0.078$$

- It helps here to visualize the "formats" and calculations with Venn diagrams.

What Does "The Probability Calculus" Say? Part II

Probability Universe (R)	Frequency Universe ($\mathcal{S}$)
	
$d = 1 - (a + b + c)$	$\delta = 1000 - (\alpha + \beta + \gamma)$

- The Venn diagram representations make it clear that we don't just have two different "formats" in which *the same information* is encoded — we have *two different sets of given information*, and *two different corresponding problems*:

Pr: Given: $a + b = .01$, $\frac{b}{a+b} = .8$, $\frac{c}{1-(a+b)} = .096$. Compute: $\frac{b}{b+c} = \Pr_R(H \mid E)$.

f: Given: $\frac{\alpha+\beta}{\#(\mathcal{S})} = \frac{10}{1000}$, $\frac{\beta}{\alpha+\beta} = \frac{8}{10}$, $\frac{\gamma}{\#(\mathcal{S})-(\alpha+\beta)} = \frac{95}{990}$. Compute: $\frac{\beta}{\beta+\gamma} = f_{\mathcal{S}}(\mathcal{C} \mid \mathcal{P})$.

We Seem to Do "Better" with the "Frequency Format"

- Gigerenzer's experiments [11] show that a far greater percentage ($\approx$ 50% *vs* $\approx$ 10%) of subjects gives the "correct" answer under the "frequency format".
- As we have seen (and as Gigerenzer [11] notes), the "frequency format" leads to a problem which is simpler, *computationally*, than the "probability format".
- Gigerenzer claims that this is only *part* of the explanation. He thinks the other part requires some kind of *evolutionary* story about the "ecological rationality" of frequencies ("frequency reasoning" was *selected for*?).
- I'd like to suggest some alterative ways of interpreting and evaluating the experimental results, and some alterative explanations of them as well. [See [21] for an excellent survey of various issues (both psychological and philosophical) that have been raised in the base rates literature over the years.]
- I begin with some dissension from two eminent philosophers of probability.

Is the "Correct" Answer *Correct*? Part I: Some Dissension . . .

- Philosophers of probability have not been so quick to characterize the "probability format" responses of the experts (or nonexperts) as incorrect.

> A probability that holds uniformly of each of a class of events because it is based in causal properties . . . cannot be altered by facts, such as chance distribution, that have no efficacy in the individual events. (L.J. Cohen [6])

> The experimental subjects studied by Kahneman and Tversky *et al* seemed to have a better grasp of the matter — even from a Bayesian point of view — than do the experimental psychologists . . . it is precisely in situations of this sort that principles of insufficient reasons are invoked . . . (Isaac Levi [22])

- I think Cohen and Levi both raise interesting objections here to the negative normative judgment implicit in claiming the "correct" answer is correct.
- It seems to me that Cohen and Levi are raising *different* objections: one from a non-Bayesian perspective, and the other from a Bayesian perspective.

Is the "Correct" Answer *Correct*? Part II: Cohen

- Following Cohen, let's look carefully at the information that is given in the "probability format" of Gigerenzer's questionnaire. We are told three things:

i. The probability of breast cancer is 1% for a woman at age 40 who participates in screening ($\mathcal{R}$).

ii. If a woman has breast cancer, the probability is 80% that she will get a positive mammogram.

iii. If a woman doesn't have breast cancer, the probability is 9.6% that she'll get a positive mamm.

- In the calculation of the "correct" answer, Bayes' Theorem is used. So, it is assumed that the "base rate" (*i.e.*, the unconditional probability) in (*i*) derives from *the same probability distribution* as the likelihoods in (*ii*) and (*iii*).
- It is interesting to note that Gigerenzer does not refer to the *reference class* $\mathcal{R}$ in his (*ii*) and (*iii*). Why not? Cohen: Because the likelihoods are *invariant* across (a wide range of) different reference classes, but the base rates are *not*.
- That is, $\Pr_R(E \mid \pm H)$ does not depend (sensitively) on R, but $\Pr_R(H)$ does.
- Cohen?: $\Pr_R(E \mid \pm H)$ is a *propensity*, but $\Pr_R(H)$ is a (mere, actual) *frequency*.

Is the "Correct" Answer *Correct*? Part II: Cohen (cont'd)

- If $\Pr^*(E \mid H)$ is interpreted as the propensity (or disposition) for the presence of breast cancer in a patient to bring about a positive test result (assuming a fixed test procedure and regime), then what in the world is its "inverse" $\Pr^*(H \mid E)$?
- What is the "propensity of a positive test result to bring about the presence of breast cancer in a patient"? This is "Humphreys' Paradox" [28], [8], [16].
- Humphreys' Paradox has lead some philosophers of probability ([9], [16]) to conclude that *propensities themselves don't obey the probability calculus*!
- Some philosophers of probability have provided ways to interpret $\Pr^*(H \mid E)$. Gillies [12] and Miller [24] both provide sensible interpretations of $\Pr^*(H \mid E)$.
- I won't go into the details of these approaches to "inverse propensities". But, the calculation in (1) which lead to the "correct" answer is no longer obviously correct (and the 70% – 80% "expert" range isn't obviously incorrect) if we interpret $\Pr_R(H \mid E)$ and $\Pr_R(E \mid H)$ as *propensities* (see also [15, page 421]).

Is the "Correct" Answer *Correct*? Part III: Levi

- If we think of the subjects as "good Bayesians", then we suppose that they have degrees of belief Pr in the salient propositions (E, H, and their logical combinations), and that these degrees of belief obey the probability calculus.
- We thus assume that, when calculating $\Pr(H \mid E)$, they will make use of *both* their likelihoods $\Pr(E \mid \pm H)$ *and* their unconditional probabilities $\Pr(\pm H)$.
- That is, their Pr must satisfy the following "odds form" of Bayes' Theorem:

(3)
$$\frac{\Pr(H \mid E)}{\Pr(\neg H \mid E)} = \frac{\Pr(E \mid H)}{\Pr(E \mid \neg H)} \cdot \frac{\Pr(H)}{\Pr(\neg H)}$$

- Since the error characteristics of the mammogram are *invariant* and *resilient* [29], a good Bayesian *should* take them on-board as *their own likelihoods*:

(4)
$$\frac{\Pr(H \mid E)}{\Pr(\neg H \mid E)} = \frac{.8}{.096} \cdot \frac{\Pr(H)}{\Pr(\neg H)} = 8.33 \cdot \frac{\Pr(H)}{\Pr(\neg H)}$$

- $\therefore \Pr(H) \in [.22, .32] \Rightarrow \Pr(H \mid E) \in [.7, .8]$, and $\Pr(H) = .5 \Rightarrow \Pr(H \mid E) = .89$. *Should* a Bayesian adopt the given base rates as *their own priors*? See [2].

Some *Explananda* and *Explanans* I

- It seems to me that there are two distinct *explananda* here:
  - (5) Why do subjects conform more closely to the "correct" answer in the "frequency format" than in the "probability format" ($\approx$ 50% *vs* $\approx$ 10%)?
  - (6) Why do even "experts" tend to give answers in the range [.7, .8] (rather than some *other* "incorrect" range) in the "probability format"?
- Gigerenzer appeals to the following two *explanans* for (5):
  - (7) The "frequency format" makes the "correct computation" *simpler*.
  - (8) Frequencies are "ecologically superior" (wtm) to probabilities.
- I suggest that the implicit frequency/propensity *ambiguities* in the "probability format" (as pointed out by Cohen/Levi) suffice [together w/(7)] to explain (5).
- (7) and (8) are irrelevant to (6). Gigerenzer [11, p. 114] does try to "identify" various "non-Bayesian" strategies, but in an *ad hoc*, disunified way, and only with small percentages of "incorrect" responses (many were "unidentified").

Some *Explananda* and *Explanans* II

- I think Cohen does provide some potential *explanans* for (5), *via* the frequency/propensity ambiguity of "probability" in the questionnaire. In the "frequency format," there is no such ambiguity (there are *only* frequencies).
- Levi (+ a little Skyrms) provides some good reasons to wonder why the Bayesian *should* conform to the "correct" answer in the "probability format". This goes *some* way toward a Bayesian explanation (or *rationale*) of (5).
- Neither Cohen nor Levi seems to have a very good explanation of (6).
  - Is Cohen suggesting that $\Pr(H \mid E)$ should be *equated* with $\Pr^*(E \mid H) = .8$? This *could* explain (6), but seems *wrong*. Perhaps a theory of "inverse propensity" might help Cohen here, but then *more information* (not in the questionnaire) [*e.g.*, $\Pr^*(H)$?] would be needed to compute $\Pr^*(H \mid E)$.
  - Levi's naive account yields $\Pr(H \mid E) = .89$, which is well outside the observed range [.7, .8]. One could *assume* that "experts" should be such that $\Pr(H) \in [.22, .32]$, but *why* (this seems like *ad hoc accommodation*)?

An Important Conflation in the Philosophical Literature I

- Carnap [3] gives an explication of $\mathfrak{c}(H, E)$: the degree to which E confirms H, in which $\mathfrak{c}(H, E) = \Pr(H \mid E)$. Popper [25, Appendix *ix] rightly points out that Carnap incorrectly conflates *degree of belief* with *degree of support*.
- Degree of *support* should be a function of the *relevance* of E to H, whereas degree of *belief* shouldn't [$\Pr(H \mid E)$ can be high even if E is *irrelevant* to H].
- Carnap [4] distinguishes "confirmation as firmness" [$\Pr(H \mid E)$] and "confirmation as increase in firmness" [$\Pr(H \mid E) - \Pr(H)$] in response to Popper's critique. And, modern Bayesian confirmation theory was born.
- In my dissertation [10], I survey the wide variety of Bayesian measures of degree of support that have been proposed and defended since 1900.
- I argue that the correct *explicatum* must be some monotonic function of the *likelihood ratio* $\frac{\Pr(E \mid H)}{\Pr(E \mid \neg H)}$. On a [-1,1] scale, $l(H, E) = \frac{\Pr(E \mid H) - \Pr(E \mid \neg H)}{\Pr(E \mid H) + \Pr(E \mid \neg H)}$. Several other authors have given independent motivation for l ([19], [13], [14], [27]).

An Important Conflation in the Philosophical Literature II

- Possible *Rationale* (Bayesian or non-Bayesian): perhaps the subjects — either confused by the different senses of "probability" suggested in the questionnaire, or sensitive to the fact that $\Pr(H \mid E)$ is not objectively determined solely by the information in the questionnaire — *should* go for the closest objectively determined "probability-like" concept: *degree of support*.
- This could provide a *rationale* (probably *not* an *explanation*) for the "expert" answers, since $l(H, E)$ is *objectively determined by the given likelihoods alone, and does not depend on priors or base rates*. In this case, we have:

$$l(H,E) = \frac{\Pr(E \mid H) - \Pr(E \mid \neg H)}{\Pr(E \mid H) + \Pr(E \mid \neg H)} = \frac{.8 - .096}{.8 + .096} = .78 \in [.7, .8] \tag{10}$$

- Other functions of $\Pr(E \mid \pm H)$ (*e.g.*, $\Pr(E \mid H)$, $\Pr(E \mid H) - \Pr(E \mid \neg H)$) are *also* on [.7,.8] ([11]), but *they* are *inadequate* measures of degree of support [10].
- See [31] and [20] for more sophisticated contemporary *psychological* models of probability judgment that are based on accumulation of evidential *support*.

References

[1] M. Bar-Hillel, *The base rate fallacy in probability judgement*, Acta Psychologica **44** (1980), 211–233.

[2] ______, *Back to base rates*, Insights in Decision Making: A Tribute to Hillel J. Einhorn (R.M. Hogarth, ed.), University of Chicago Press, 1990, pp. 200–216.

[3] R. Carnap, *Logical Foundations of Probability*, 1 ed., University of Chicago Press, 1950.

[4] ______, *Logical foundations of probability*, 2 ed., University of Chicago Press, 1962.

[5] W. Casscells, A. Schoenberger, and T. Grayboys, *Interpretation by physicians of clinical laboratory results*, New England Journal of Medicine **299** (1978), 999–1001.

[6] L.J. Cohen, *Can human irrationality be experimentally demonstrated*, Behavioral and Brain Sciences **4** (1981), 317–370.

[7] D. Eddy, *Probabilistic reasoning in clinical medicine: Problems and opportunities*, in [17].

[8] J. Fetzer, *Scientific knowledge: Causation, explanation, and corroboration*, D. Reidel, 1981, *Boston Studies in Philosophy of Science*, Volume 69.

[9] ______, *Probabilistic explanations*, PSA 2 (1982), 194–207.

[10] B. Fitelson, *Studies in Bayesian confirmation theory*, Ph.D. thesis, University of Wisconsin–Madison (Philosophy), 2001, `http://fitelson.org/thesis.pdf`.

[11] G. Gigerenzer, *Adaptive thinking*, Oxford University Press, 2000.

[12] D. Gillies, *Philosophical theories of probability*, Routledge, 2000.

[13] I.J. Good, *Weight of evidence: A brief survey*, Bayesian Statistics, 2 (J Bernardo, M. DeGroot, D. Lindley, and A. Smith, eds.), North-Holland, 1985, pp. 249–269.

[14] D. Heckerman, *An axiomatic framework for belief updates*, Uncertainty in Artificial Intelligence 2 (L. Kanal and J. Lemmer, eds.), Elsevier Science, 1988, pp. 11–22.

[15] C. Howson and P. Urbach, *Scientific reasoning: The Bayesian approach*, Open Court, 1993.

[16] P. Humphreys, *Why propensities cannot be probabilities*, The Philosophical Review **94** (1985), 557–570.

[17] D. Kahneman, P. Slovic, and A. Tversky (eds.), *Judgment under uncertainty: Heuristics and biases*, Cambridge University Press, 1982.

[18] D. Kahneman and A. Tversky, *On the psychology of prediction*, Psychological Review **80** (1973), 237–251.

[19] J. Kemeny and P. Oppenheim, *Degrees of factual support*, Philosophy of Science **19** (1952), 307–324.

[20] D. Koehler, C. White, and R. Grondin, *An evidential support accumulation model of subjective probability*, *Cognitive Psychology*, in press, 2002.

[21] J. J. Koehler, *The base rate fallacy reconsidered: Normative, descriptive and methodological challenges*, Behavioral and Brain Sciences **19** (1996), 1–53.

[22] I. Levi, *Should Bayesians sometimes neglect base rates?*, Behavioral and Brain Sciences **4** (1981), 342–343.

[23] D. Lyon and P. Slovic, *Dominance of accuracy information and neglect of base rates in probability estimation*, Acta Psychologica **40** (1976), 287–298.

[24] D. Miller, *Critical rationalism: A restatement and defence*, Open Court, 1994.

[25] K. Popper, *The logic of scientific discovery*, Routledge, 1992, Translated from the German, Reprint of the 1980 English translation.

[26] R. Rosenkrantz, *Foundations and applications of inductive probability*, Ridgeview, 1981.

[27] Richard M. Royall, *Statistical evidence: A likelihood paradigm*, Chapman & Hall, 1997.

[28] W. Salmon, *Propensities: A discussion review of D.H. Mellor The Matter of Chance*, Erkenntnis **14** (1979), 183–216.

[29] B. Skyrms, *Pragmatics and empiricism*, Yale University Press, 1984.

[30] A. Tversky and D. Kahneman, *Causal schemas in judgements under uncertainty*, in [17].

[31] A. Tversky and D. Koehler, *Support theory: A nonextensional representation of probability judgment*, Bayesian statistics, 5 (J. Bernardo, J. Berger, A. Dawid, and A. Smith, eds.), Oxford Univ. Press, 1996, pp. 445–460.