- **Recall (Lecture XVI)** ***Wesley Salmon: "Rationality in Science or Tom Bayes Meets Tom Kuhn"***

> **Salmon's project:** Can Bayes' Thm. (interpreted in terms of theory-confirmation) add that 'extra element' to the process of theory-choice (outside observation/experimentation) and hypothetico-deductive methods that Kuhn was alluding to? Most important, will Bayesianism prove sufficient to filter out adverse consequences of some *contextual values* that would undercut scientific rationality and objectivity, while preserving and adding an element of systematicity to the best that some of the *cognitive values* of Kuhn's criteria could offer?

Salmon is cautiously optimistic: "[Bayes'] algorithms are trivial; **what is important is the scientific judgment involved in assessing the probabilities that are fed into the equations. The algorithms give frameworks in terms of which to understand the role of the sort of judgment upon which Kuhn rightly placed great emphasis…**I do not for a moment maintain that complete consensus would be in the offing even if both camps were to buy the Bayesian line I have been peddling. But I do hope that some areas of misunderstanding have been clarified." (579-580)

- Swamping/Washing Out the Prior Probabilities: Recall Salmon's trick:

For any two theories T_i , T_j : (where: $1 \le i, j < n$ [1] , $i \ne j$)

$$\frac{P(T_i \mid E \cap B)}{P(T_j \mid E \cap B)} = \frac{P(E \mid T_i \cap B)P(T_i \mid B)}{P(E \mid B)} \Bigg/ \frac{P(E \mid T_j \cap B)P(T_j \mid B)}{P(E \mid B)} = \frac{P(E \mid T_i \cap B)P(T_i \mid B)}{P(E \mid T_j \cap B)P(T_j \mid B)}$$

…which reduces the (admittedly hopeless task) of evaluating a posterior probability of some theory in the light of comparison with *all* alternatives, including the 'catchall'[2] to the case of the *comparative evaluation* of two (or more) *bona fide* (non catchall) theories. The evidential value term or expectedness simply cancels in the denominator terms. The resulting comparative evaluation scenario is closer to the actual scenario depicted by Kuhn, concerning values and theory-choice.

Hence (deriving Eqn.7), one has grounds to prefer theory T_i over T_j in terms of their posterior probabilities whenever (obviously):

[1] The strict inequality on the right hand side is deliberate, since neither of these two theories can be the *catchall* theory: $T_n = \sim(T_1 \cup T_2 \cup \ldots \cup T_{n-1}) = (\sim T_1 \cap \sim T_2 \cap \ldots \cap \sim T_{n-1})$

[2] "To try to evaluate the likelihood of the catchall involves, it seems to me, an attempt to guess at the future history of science. That is something we [obviously] cannot do with any reliability." (569)

$$\frac{P(T_i \mid E \cap B)}{P(T_j \mid E \cap B)} > 1$$

which of course holds whenever: $\frac{P(E \mid T_i \cap B)P(T_i \mid B)}{P(E \mid T_j \cap B)P(T_j \mid B)} > 1 \Rightarrow \frac{P(E \mid T_i \cap B)}{P(E \mid T_j \cap B)} > \frac{P(T_j \mid B)}{P(T_i \mid B)}$

…or in other words, when the **likelihood ratio of T_i versus T_j** is greater than the **prior probability of T_j versus T_i.** Hence, for instance, the rather contentious issue regarding the 'washing out of the prior probabilities' (which supposedly guarantees convergence to a reliable answer for the posterior probability) is reduced to the special case when one has no prior preference of one theory over the other, i.e. when $\frac{P(T_j \mid B)}{P(T_i \mid B)} = 1$. In that case, **the likelihood ratio 'washes out' the prior probability ratio**, (in a trivial sense) since the latter =1, so the choice of theory is determined by the former.

- **Coping with Anomalies (Scenarios/ Auxiliary Hypotheses)**

In the potentially incommensurable case[3] of anomalies (i.e. some item(s) of evidence E', E'' respectively confronting T_i or T_j such that: $P(E' \mid T_i \cap B) \cong 0$ or $P(E'' \mid T_i \cap B) \cong 0$, i.e. rendering their likelihoods virtually zero) so Salmon argues the Bayesian framework can accommodate this issue. **Plausible scenarios (or *auxiliary hypotheses*)** $\boldsymbol{A_i}$ **,** $\boldsymbol{A_j}$ are adjoined to the theories T_i and T_j rendering their likelihoods far higher, "possibly unity."[4] (571))[5]:

$$\frac{P(T_i \cap A_i \mid E \cap B)}{P(T_j \cap A_j \mid E \cap B)} = \frac{P(E \mid T_i \cap A_i \cap B)P(T_i \cap A_i \mid B)}{P(E \mid T_j \cap A_j \cap B)P(T_j \cap A_j \mid B)}$$

"We have…handled the central issue raised by Kuhn…what is the basis for preference between two theories." (573) (though "these formulas provide no evaluations of individual theories; they furnish only comparative evaluations." (574)

- **Salmon's Qualifications**

Recall Kuhn's criteria: simplicity, accuracy, scope, fecundity, consistency. They can be categorized according to the following kinds of virtues: informational, confirmational, economic.[6] **"[O]ur use of Bayes's theorem is germane only to the confirmational virtues." (574)** "Kuhn's criteria [are]…intended to show how adequately they can can be understood in a Bayesian framework—insofar as they are germane to confirmation. **If**

[3] Anomalies can be an issue whether or not one buys into a notion of paradigm/theoretical framework.

[4] Recall some of the examples: adjoining the scenario of the *aether* hypothesis to account for the theory of transverse light waves, to overcome the anomaly of empty space (which at the time the theory of wave mechanics would deliver a zero likelihood for waves to propagate in, i.e. waves we required to propagate in *something*.) Another example is the oft-cited hypothesis that the stars we far more distant from the solar system, to account for the lack of observed stellar parallax motion, in the case of the Copernican theory.

[5] For those mathematically motivated: consider this adjustment as depicted in the above expression (or Eqn. (8), 573) as a means to avoid the $^0/_0$ indeterminate form, since as mentioned above anomalies can evince zero likelihoods for T_i as well as T_j.

[6] Simplicity ("broadly construed"-576) and consistency are confirmational, while fecundity and scope are informational. Accuracy can be construed both in terms of informational as well as economic virtues.

it is sound, we have constructed a fairly substantial bridge connecting Kuhn's views on theory choice with those of the logical empiricists—at least, for those who find in Bayes' theorem a suitable schema for characterizing the confirmation of hypotheses and theories." (577) There are also deductive as well as inductive relations among theories. The latter are characterized by analogies.
Kuhn of course mentioned also that such virtues can conflict: one person can select T_i because of its scope while another may select T_j because of its simplicity. **Karl Popper, of course, placed constraints on scope. (575)**

Note on consistency: "The discovery of an internal inconsistency has a distinctly adverse effect on the prior probability of that theory, to wit, it must go straight to zero." (576) **Counterexample: Classical electrondynamics.[7]**

Objectivity[8] (578)	
static	The removal of inconsistencies or contradictions
kinematic	*Constraining* one's belief(s)[9] according to the strictures of Bayes's formula. (I.e., Bayeasian conditionalization.)
dynamic	Introducing an *objective interpretation* to the prior probabilities in Bayes' formula (by way of *a frequentist approach.*[10]) Note the obvious appeal to actuality here (as mentioned in n. 9 below)

"[Bayes'] algorithms are trivial; **what is important is the scientific judgment involved in assessing the probabilities that are fed into the equations."** (579)

Clark Glymour : "Why I am Not a Bayesian"

"The aim of confirmation theory ought not to be simply to provide precise replacements for informal methodological notions, that is, explications of them. **It ought to do more; in particular, confirmation theory ought to *explain* both methodological truism and particular judgments that have occurred within the history of science…I mean at least that confirmation theory ought to provide a rationale for methodological truisms [and]…reveal some systematic connections among them…without arbitrary or question-begging**

---

[7] Recall **Lecture I** : Mathias Frisch's study of the theory revealing its deep internal inconsistencies viz. the issue of discrete charge distributions (in that case Maxwell's equations, the Lorentz force law, and energy conservation form an inconsistent set). No practicing eelectrical engineer nor for that matter any experimental physicist heavily relying on the theory would agree with Salmon's assessment concerning prior probabilities.

[8] The classifications: *static, kinematic, dynamic* come from physics. The notion 'static' is somewhat obvious. **However, there's a subtle but important distinction between kinematic and dynamic.** Kinematics is the *mere decription of motion* or *possible classes of motion* (where 'possible' is defined in terms of what theory's laws allow). Whereas *dynamics* is described in terms of the *particalur cause(s) of motion of an actual system as it evolves in time.* For example: consider the elementary example of projectile motion (neglecting air resistance). Kinematic analysis deals with the *class of curves* described by such motion (namely parabolae). Whereas *dynamic analysis* focus on the *causes of the actual motion* of such a system exhibiting such motion (which usually reduces to talk of initial conditions & a parameterized description of the motion of the system along the parabolic trajectory, with *t* [time] being the typical default parameter.) In modal terms, then, *kinematics* is naturally ascribed to possibility the way dynamics is to actuality.

[9] See note 8 above.

[10] Recall **Lecture XV**.

assumptions to reveal particular historical judgments as in conformity with its principles." (CC1998 584-585)

- **According to Glymour, Bayesian philosophers of science seem ultimately to be guilty of committing a *category mistake* [11] betweem normative and descriptive accounts:**

"[P]articular *inferences* can almost always be brought into accord with the Bayesian scheme by assigning degrees of belief more or less ad hoc, *but we learn nothing from this agreement. What we want is an explanation of scientific argument; what the Bayesians give us is a theory of learning, indeed a theory of personal learning. But arguments are more or less impersonal…*To ascribe to me degrees of belief that make my slide to my premises to my conclusion a plausible one fails to explain anything not only because the ascription may be arbitrary, but also because, even if it is a correct assignment of my degrees of belief, it does not explain why what I am doing is *arguing*--…that is, what I say should have the least influence on others, or why I might hope that it should. *"* (592, some italics added)

-For one thing, though a relatively minor point, Glymour disputes the analogy of Carnap to Frege viz. presumed "breakthrough" in the fields of deductive and inductive logic, respectively.[12] "[E]xplicitly, probability is a distinctly minor note in the history of scientific achievement…my own inclination is to believe that the interest such investigations have stems more from the insights they obtain into syntactic versions of structural connections among evidence and hypotheses than to the probability measures they mesh with these insights." (586)

-(**Recall problem of fixing a sample space[13]) Problem of fixing reference class/ prior prabilities:** "Frequency interpretations suppose that for each hypothesis to be assessed there is an appropriate reference class of hypotheses to which, and the prior probability of the hypothesis [as well as for singular and plural evidence-statement(s)] is the frequency of true hypotheses in the reference class…**The matter of how such reference classes are to be determined…so that [for example] the frequencies do not come out to be zero[14], is a question that has only been**

---

[11] Such a notion was originally developed in systematic detail by analytic philosopher of language Gilbert Ryle: Committing category mistakes (i.e. most commonly-occurring involve the positing of 'philosophical problems' arising from a basic confusion of basic categories: e.g., epistemology with ontology or methodology, etc.)

[12] Before Gottlob Frege, though mathematician and logician George Boole in the 1840s developed an algebra which characterized the Aristotelean logic of statements and propositions in some systematically rigorous manner (i.e. the *Boolean algebra*), it was only Frege who developed a fecund (recall Kuhn's criterion) research tradition of mathematical logic which (among other things) lead to a axiomatic precise formalization of mathematical induction, *n*-th order predicate logic. Despite his astonishing achievements (only truly recognized and appreciated posthumously) Frege's goals were relatively modest: he "simply" wanted to show that Kant was wrong, i.e. that mathematics was not based on any extra 'synthetic/intuitive' (constructive) faculty, but could be fully reduced to *a priori* or analytic logical/axiomatic principles *provided such principles were descriptive and therefore rich enough.* (Hence Frege's *radical* extension of propositional logic). Most contemporary ('neo-Fregean') scholars argue that Russell's criticisms of Frege's Axiom V, presumably revealing a basic inconsistency and paradox in his research programme, were superficial and premature. Frege aside, Carnap's treatise on probability theory was an attempt to secure axiomatic and *a priori* foundations to inductive forms of reasoning in a manner analogous to Frege's overhaul of deductive reasoning. Glymour expresses profound skepticism at the latter's attempt(s): **"[P]robabilistic analyses remain at too great a distance from the history of scientific practice to be really informative about that practice."** (585)

[13] From notes on Hawthorne

[14] For instance, in the same page Glymour asks what the Bayesian has to say about the case concerning useful hypotheses that are practiced across scientific communities, that turn out to be literally false (hence

touched on by frequentist writers…the history of scientific argument must turn out to be largely a history of fanciful guesses. " (586)

Glymour's three main critical theses contra Bayesian approaches in the philosophy of science (587):

1. Arguments for the *a priori* fixation of belief and rationality according to Bayesian principles "remain unconvincing."[15]
2. Only *some* of the methodological principles and truisms most commonly accepted in scientific practice[16] can be leant a Bayesian account.
3. "[T]here are [also] elementary but perfectly common features of the relation of theory and evidence that the Bayesian scheme cannot capture without serious—and perhaps not very plausible—revision." (587)

(Re. Point 1 above): Glymour questions the very assumption that –though it's plausible we have *kinds* of belief—it's less clear our beliefs occur in 'grades' in such a way that can be captured by the axioms of probability.

According to Glymour, the latter assumptions that belief occurs in grades (which are ultimately characterized by the axioms of probability) has to do iwh the 'zero sum' scenario of betting:

An ideally rational agent, when s/he is informed s/he may receive $U + V$ $ in return for the maximum amount $U$$ s/he invested in a bet that proposition p is or isn't true. If however s/he loses the bet, s/he will receive 0$, and hence must write off the $U$$ invested, i.e. s/he must declare the $U$$ investment a loss. Then[17]:

$$VP(p \mid B) + -UP(\sim p \mid B) = 0 = VP(p \mid B) - U(1 - P(p \mid B))$$
$$\Rightarrow (V + U)P(p \mid B) = U \Rightarrow P(p \mid B) = \frac{U}{U + V}$$

Hence, so the Bayesian claims, degree of belief associated with the above objective probability/ betting odds quotient. **But this assumption is tendentious!** "The subject may not believe that the bet will be paid off if [s/]he wins, or [s/]he may doubt that it is clear what constitutes winning …[and losing…**And the very fact that [s/]he is offered a wager on [*p*] may somehow change his degree of belief in [*p*] ."** (589)

"[W]hy should we think that, for rationality, one's degrees of belief must satisfy the axioms of probability, and why should we think that, again for rationality, changes in belief ought to proceed by conditionalization?…So the [above] betting quotient determines a degree of

---

their prior probabilities = 0). The Bayesian would want to jettison such theories, despite their potentially deeply useful nature (as evidenced by their widespread use).

[15] According to Glymour, Bayesians concede this point as well.

[16] E.g. preference for some kinds of evidence, disdain for *ad hoc* principles (as often embodied by 'deoccamized' theories, i.e. theories which don't adhere to Ockham's Razor [an example of Kuhn's **simplicity** criterion] which states that given two *prima facie* correct demonstrations/arguments, select the one with the fewer number of steps and auxiliary assumptions. In addition, (recall Longino) knowing what makes some evidence *relevant* for a theory.

[17] I changed the article's notation slightly in the case of the probability, to depict it in a more explicitly conditional fashion: $P(p \mid B)$ is the probability that proposition p is true, given certain background conditions B.

belief…in the *role* of a probability. **But why should…the degrees of belief that play this role, be probabilities?"** (589-590)

For instance (recalling Longino) one could have contextual values such that "[such] values [are] other than the value we place on the stakes and these other values may enter our determination whether or not to gamble; and we may not have adopted a policy of acting so as to maximize our expected gain or our expected utility…we may save ourselves…by refusing to make certain wagers, or combinations of wagers, even though we judge the odds to be in our favor." (590)

- Bayesians argue that the connection between what *is inferred* versus what *is the case* can be established via long-run convergence principles, via Savage's Theorem[18]:

-Let $\{T_i\}$ be a set of m mutually exclusive (i.e. $T_i \cap T_j = \varnothing$, for all $1 \leq i, j \leq m.\ i \neq j$) and exhaustive[19] theories/hypotheses, with their prior probabilities $P(T_i \mid B)$
-Consider a finite sequence of random variables[20] $X_r = \langle x_1, x_2, \ldots, x_N \rangle$ with conditional distribution $P(x_k \in X_r \mid T_j \cap B) \equiv f(x_k \mid T_j)$ [21] such that for any T_i and T_j there exists at least one $x_l \in X_r$ such that: $f(x_l \mid T_i) \neq f(x_l \mid T_j)$, i.e. **no two hypotheses/theories determine the same likelihood distribution**. Let $X_{r,n}$ be a subsequence of X_r containing X_r's first n elements (where $1 \leq n \leq N$), i.e.: $X_{r,n} = \langle x_1, x_2, \ldots, x_n \rangle \subseteq X_r = \langle x_1, x_2, \ldots, x_N \rangle$.
-Then prior to an observation of all the elements in $X_{r,n} = \langle x_1, x_2, \ldots, x_n \rangle$, for any $0 < \alpha < 1$, and for any $x \in X_{r,n}$:[22]

$$P(P(x > \alpha) \mid B) = P[(P(T_1 \mid x) > \alpha) \mid T_1 \cap B]P(T_1 \mid B) + P[(P(T_2 \mid x) > \alpha) \mid T_2 \cap B]P(T_2 \mid B) + \ldots + P[(P(T_m \mid x) > \alpha) \mid T_m \cap B]P(T_m \mid B)$$

Then: As $n \rightarrow \infty$ [23],

$$P(P(x > \alpha) \mid B) = P[(P(T_1 \mid x) > \alpha) \mid T_1 \cap B]P(T_1 \mid B) + P[(P(T_2 \mid x) > \alpha) \mid T_2 \cap B]P(T_2 \mid B) + \ldots + P[(P(T_m \mid x) > \alpha) \mid T_m \cap B]P(T_m \mid B) \rightarrow 1.$$

- (Re. Points 1 and 2 above): **Glymour accuses the Bayesian as begging the question, should s/he cite Savage's Theorem as a reason as a long-term convergence principle:**

"The result [of Savage's Theorem (ST)] involves second-order probabilities,[24] but these too, according to personalists, are degrees of belief. So what has been shown seems to be this: In the limit as n approaches infinity an ideally rational Bayesian has degree of belief 1 that an ideally rational Bayesian (with degrees of belief as in the theorem) has degree of belief [as stated in the conclusion of ST]…**The theorem does not tell us that in the limit any rational**

[18] I have changed the notation from Glymour's here, for purposes of greater clarity.
[19] For this condition to hold, one of the B's *must* be the catchall, i.e. suppose (with no loss of generality) it's the last one: $T_m = \sim(T_1 \cup T_2 \cup \ldots \cup T_{m-1}) = (\sim T_1 \cap \sim T_2 \cap \ldots \cap \sim T_{m-1})$
[20] One can think of such a sequence as a set of measurement outcomes. (590)
[21] Where (obviously) $1 \leq k \leq N$ and f is some function satisfying the axioms of probability (recall Hawthorne's seven axioms in the case of inductive support functions.)
[22] Note that this partitioning is the same scheme as when the denominator term was partitioned in the Bayes' formula in **Lecture XIV** and **Lecture XV**. For instance:

$$P(E \mid B) = P(E \mid T_1 \cap B)P(T_1 \mid B) + P(E \mid T_2 \cap B)P(T_2 \mid B) + \ldots + P(E \mid T_n \cap B)P(T_n \mid B)$$

[23] Which obviously entails that $N \rightarrow \infty$. This is just a case when arbitrarily many measured quantities are registered
[24] I.e., probabilities of probabilities

Bayesian will assign probability 1 to the true hypothesis and probability 0 to the rest; it only tells us that rational Bayesians are certain that [s/]he will." (591)

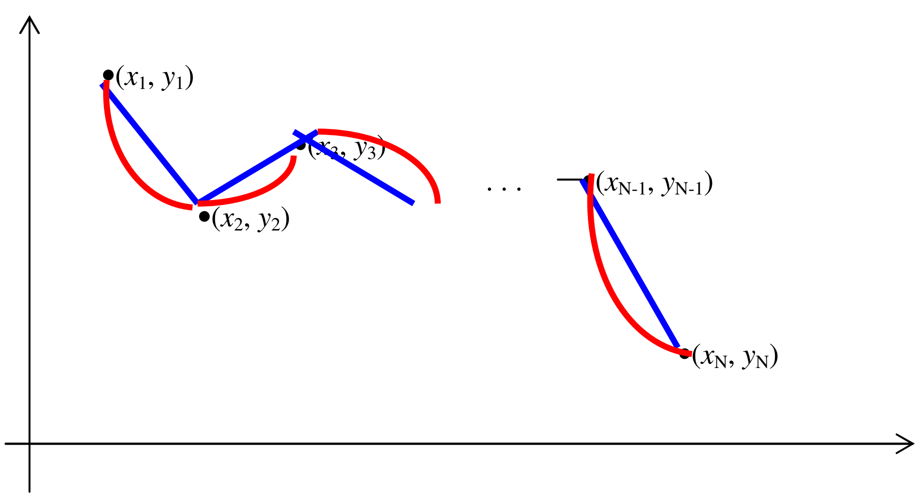
…Or stated in more general terms (recalling the 'Category mistake' between descriptive and normative accounts of reasoning alluded to above):

"[M]ore hopefully, **Bayesians may suggest we give arguments exactly because there are general principles restricting belief**, principles that are widely subscribed to, and in giving arguments we are attempting to show that, supposing our audience has certain beliefs, they must in view of these principles have other beliefs**…[But] [w]hat is controversial is that the general principles required for argument can best be understood as conditions restricting prior probabilities in a Bayesian framework.** *Sometimes* they can, perhaps, but when **I think that arguments turn on relating evidence to theory, it is very difficult to explicate them in a plausible way within the Bayesian framework."** (592, italics added)

Note2: Is Glymour setting the bar a little too high here? Recall (**Lecture II**) that logic is obht a descriptive and a normative enterprise. Should one consider Bayesians as advocating some garden-variety "inductive logic" (recall Hawthorne's article) are such descriptivist notions that tepid, in terms of explanatory value? Much depends on what one requires from an "explanation," as we'll see in the ensuing section of this course.

- (Re. Point 3 above): **Relation of theory to evidence (interpolation schemes)**

Suppose one had a finite collection of data: $\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$. Then *ceteris paribus* one prefers an interpolation scheme that introduces the least number of new parameters, which translates to a *linear* scheme (should one prefer a point-wise scheme. Otherwise one adopts a *best-fitting* straight line, via the procedure of *Least Squares Interpolation*):



Example: (linear versus quadratic) $y_k^1(x) = m_k x + b_k$ (for $x_{k-1} \leq x \leq x_k$) where the two parameters m_k b_k are easily obtained through the *point-slope* routine)

On the other hand $y_k^q(x) = A_k x^2 + B_k x + C_k$ (for $x_{k-1} \leq x \leq x_k$) aside from introducing three parameters, also require more sophisticated numerical analysis algorithms for determination.

"If one's equation [in the general, best-fitting curve case] fits the data *too* well, then the equation has too many terms and too many arbitrary parameters; and if the equation does not fit the data well enough, then one has not included enough terms and parameters in the equation. The whole business depends…on the ordering of prior probabilities." (594)

Note 3: The excerpt from Hesse (595) in which she mentions 'gruified' hypotheses is a reference to Nelson Goodman's 'New Riddle of Induction.' One is likely to conclude for example that 'all emeralds are green' but suppose they contain the 'grue' property described as green before time *T* and blue after such *T*. Aside from violating an (Salmon's) economic virtues, such 'warped predicates' argue metaphysicians aren't bona fide in the sense that there's no clear correspondence with such (logical) notions and a corresponding metaphysical simple property. This is also true for predicates which are formed through negation[25] as well as disjunction.[26]

- Glymour adopts a *reductio ad absurdum* by exploiting an essential indeterminacy claim, in response to Hesse's 'clustering' constraint applied to prior probabilities viz. curve fitting:

  "For surely, if $[y_k^{l,1}(x) = m_k x + b_k]$[27] is a legitimate predicate, then so is $[y_k^{l,2}(x) = M_k x^2 + b_k]$ for any definite values of $[m_k$ and $M_k]$[28]. Now Hesse's first two data points can be equally well described by $(x_1, m_1 x_1 + b_k)$ and $(x_2, M_2 x_2^2 + b_k)$…**so by the clustering postulate the probability that the third point satisfies the linear expression be greater than one-half, and the probability that the third point satisfies the linear expression must also be greater than one-half, which is impossible.**" (596)

Note 4: Glymour is invoking implicitly in his arguments central indeterminacy we will examine in greater detail in this course: the Quine-Duhem thesis (QDT), which states that theories are always underdetermined by their evidence. In this case, by the lights of the QDT, one can invoke "trivial additional assumptions" (in the $(m_k, x) \rightarrow (M_k, x^2)$ maneuver.)

- **Glymour's criticism of Rosenkranz (and Salmon): Criticism of Static and Bayesian Kinematics**

  "The Bayesian account of confirmation makes it impossible for a piece of evidence to give us more total credence in a theory than in its observational consequences. The Bayesian way of setting things up is a natural one, but it is not inevitable, and wherever a distinction between theory and evidence is plausible, it leads to trouble." (598)

  "[A]ny attempt to found an account of the [Bayesian] case on…facts alone is simply an attempt at a hypothetico-deductive account. The problem is reduced to one already unsolved. What is needed to provide a genuine Bayesian explanation of the case in

---

25 Recall Russell's 'set of all things that is not a teacup'

26 Consider the predicate: 'tall' (if human) or 'flat and round' (if a 2D picture). Obviously no single elementary property corresponds to such a disjunction.

27 I changed the notation here to correspond with the above example

28 Assuming, of course, that $x_i \geq 0$ for all i. I.e., that the data are all non-negative.

question…is a *general* principle restricting conditional probabilities and having the effect that the distinctions about the bearing of evidence…do result." (599)

- Critique of kinematics (old evidence confirming new theory)

The historical record indicates that old evidence confirms new theories on a rather routine basis (example: the 1915 observed anomaly [to Newtonian celestial mechanical paradigm] of the perihelion shift of Mercury aided to the retrospective confirmation of General Relativity insofar as motivating the famous [and costly] expedition in 1919 to Brazil to photograph stars during a solar eclipse, for the confirmation of the gravity lens effect)

Unfortunately Bayes' algorithm cannot accommodate this retroactive confirmation!

Bayes's Formula: $$P_t(T \mid E \cap B) = \frac{P_t(E \mid T \cap B)P_t(T \mid B)}{P_t(E \mid B)}$$

(indexed with respect to time, i.e. the t subscript describes the predicate: "known at time t.")

However, since the evidence E was already known, then: $P_t(E \mid B) = 1 \Rightarrow P_t(E \mid T \cap B) = 1$ (i.e. the likelihood of E given T is also unity). Hence the above reduces to:

$$P_t(T \mid E \cap B) = P_t(T \mid B)$$

"The conditional probability of T on [E] is therefore the same as the prior probability of T; ***E* cannot constitute new evidence for *T* in virtue of the positive relevance condition nor in virtue of the likelihood of *E* on *T*."** (600)

"[I]t is beside the point that an ideal Bayesian would never face a novel theory, for the idea of Bayesian confirmation theory is to explain scientific inference and argument by means of the assumption that good scientists are, about science at least, approximately ideal Bayesians, and we have before us a feature of scientific argument that seems incompatible with that assumption." (600)

Response: Introduce counterfactual degrees of belief? "The problem is to fill in the blanks in such a way that it is both plausible that we have the needed counterfactual degrees of belief and that they do serve to determine how old evidence bears on new theory…We cannot merely throw [E] and whatever entails E out of the body of accepted beliefs; we need some rule for determining a counterfactual degree of belief in E and a counterfactual likelihood of E on T." (601)

Problem: Applying Counterfactuals to Bayes' Formula results in incoherence!

Consider theory T and a set of mutually exclsive rivals $\{T_1, T_2, \ldots, T_k\}$

Recall: $P(E|B) = P(E|T_1 \cap B)P(T_1|B) + P(E|T_2 \cap B)P(T_2|B) + \ldots + P(E|T_k \cap B)P(T_k|B) + P(E|T_{k+1} \cap B)P(T_{k+1}|B)$

Where T_{k+1} is the catchall $= \sim(T_1 \cup T_2 \cup \ldots \cup T_k) = \sim T_1 \cap \sim T_2 \cap \ldots \cap \sim T_k$

But (recall **<u>Lecture XVI</u>**) the problem of the catch-all looms large when evaluating counterfactual evidence claims (Recall Salmon's remark concerning "anticipation of all future history of physics.") Ignoring the catchall term means that the rest of the prior probabilities (of the competing theories to *T*) used to directly evaluate $P(T \mid E \cap B)$ in Bayes' formula:

$$P(T \mid E \cap B) = \frac{P(E \mid T \cap B)P(T \mid B)}{P(E \mid B)}$$

will prove incoherent unless $P(E \mid T \cap B) = 0$ or $P(T \mid B) = 0$

Other algebraic combinations and possiblilites only yield up similarly incoherent results. (602)

- Consider Jeffrey's attempt to solve the kinematic problem of old evidence via adjoining to *P* a degree of belief one would have had in *E* in some historical period (denoted by *H*(*E*)):

  $P^*(T) = H(E)P(T \mid E) + (1 - H(E))P(T \mid {\sim}E)$

  Jeffrey showed that P^* satisfies the axioms of a probability function. Conditionalizing on P^* is the actual degree of belief function: $P'(T) = P^*(T \mid E)$.

Glymour points out the implausibility of this boot-strapping technique: "It is not obvious that there are, for each of us, degrees of belief we personally would have had in some historical period." (603)

"**[I]t is implausible indeed that such a historical Bayesianism…is an accurate account of the principles by which scientific judgments or confirmations are made. For if it were, then we should have to condemn a great mass of scientific judgments on the grounds that those making them had not studied the history of science with sufficient closeness to make a judgment as to what their degrees of belief would have been in relevant historical periods." (603)**

Granted, in the rather artificial case when $T \vdash E$ (when *T* deductively entails *E*) one can speak of old evidence confirming a new theory insofar as:

$$P(T \mid E \cap B \cap (T \vdash E)) > P(T \mid E \cap B)$$

"[W]hat matters is the discovery of a certain logical or structural connection between a piece of evidence and piece of theory…[However] [w]hat I do not believe is that the relation that matters is simply the entailment relation …exactly [because of] the reasons why the H-D account…is inaccurate**; but the suggestion is at least correct in sensing that our judgment of the relevance of evidence to theory depends on the perception of a structural connection between the two, and that degree of belief is, at best, epiphenomenal**…I **think [my arguments] do at least strongly suggest that there must be relations between evidence and hypotheses that are important to scientific argument and confirmation but to which the Bayesian scheme has not yet penetrated.**" (604-605)