Notes for Week 1 of *Confirmation*

08/29/07

Branden Fitelson

1 Administrative Stuff

At the outset, I should say something about *enrollment*. Oringinally, I was planning to teach only one course on induction and probability this year (this one). As a result, I encouraged undergraduates to enroll in the seminar. Since then, I have decided to teach an undergraduate course next term on probability and induction (PHIL 148). That course will be more geared to undergraduates. As such, I want to encourage (most) undergraduates here to attend that class instead of this one. We can still accommodate a few (diehard) undergrads in the seminar (but far fewer than are enrolled, I'm afraid). Basically, I think this course will work best as a small-ish, graduate seminar. I apologize to (most) undergrads here for any inconvenience. I will collect index cards from undergrads today, and decide who to admit before next week.

The course website contains everything you need to know (and more) about this seminar. For instance, my seminar notes (including these) will be posted there prior to each meeting. Moreover, the syllabus page for the course (an early snapshot of which which I'm distributing today as a handout) explains the basic structure and content of the course. It also includes all the readings for the course (required and optional), as well as some useful overviews of the (philosophical) material we'll be discussing. I'll spend a few minutes today going over the syllabus. But, keep in mind that the schedule of readings (and my notes) will be updated as we go along. Thus, it is important to stay tuned to the course website as the semester unfolds.

2 A Broad Overview of the Plan for the Seminar

As the syllabus indicates, the course has two main parts:

- **Part I**: **Logical & epistemological aspects of confirmation theory (mainly) between 1920 and 1960**. This will be the main focus of the seminar. It will actually begin with two weeks of "Pre-History". In week 1, we will read a nice article by Milton called "Induction Before Hume", which traces some of the historical developments leading up to to Hume's *Treatise* (and *Enquiry*). Milton's piece is most useful (for us) because it sets the stage for what I will call "Keynesian" readings of Hume (which take Hume to be concerned mainly with *induction* — as opposed to *causation* — in his infamous skeptical and positive writings). After Milton, we will read Keynes and Stroud — two eloquent "Keynesian" readers of Hume. Once we've got the "Humean" background in place, we'll move on to reading various subsequent philosophers of inductive logic (and inductive inference), including Nicod, Hosiasson–Lindenbaum, Hempel, Carnap, and Goodman (among others). After Goodman, confirmation theory took a decidedly subjective turn. We will spend one week (after the "Hume → Goodman trajectory") discussing subjective Bayesian confirmation theory (and its discontents). The main issues I want to bring out in Part I involve *the relationship between inductive logic and inductive epistemology*. I'll say a little bit more about this below in section 3, when I do some "theoretical stage-setting" for the seminar.

- **Part II**: **Cognitive-scientific applications of confirmation theory (mainly) between 1980 and 2007**. This part of the course will only take up 2 weeks (out of 13), but it's a very interesting recent area of research involving confirmation theory. Unlike Part I, Part II won't really be about confirmation theory *qua logical/epistemological theory **per se***. But, the issues discussed in Part I will be relevant to understanding the theoretical and normative underpinnings of this recent work in cognitive science.[1] In Part II, we will focus on two well-known problems from the contemporary cognitive science literature:

[1] Interestingly, we can also think of many of the issues raised in Part II as having "descended from Hume". Hume's *Treatise* (and *Enquiry*) contain a very subtle and complex mixture of logical, epistemological, *and psychological* considerations. Recent work on Hume (some of which is included in the further readings for the "pre-history" of Part I, to be discussed next week) is starting to do a pretty good job of disentangling these components. I won't be talking much about Hume after the first couple weeks of the seminar (I don't even have the salient Hume texts as required readings — but they are linked as further readings early on in the seminar). But, those of you who are interested in Hume should watch for Humean influences all the way through the seminar.

The Wason Selection Task and *The Conjunction Fallacy*. As it turns out, various philosophers and cognitive scientists (including me) have been applying confirmation-theoretic ideas to these problems. Here, we'll see that confirmation theory (in the spirit of the pioneers of the field, studied in Part I) *may* turn out be useful for understanding how people *actually* think (who knew?).

3 Some Theoretical Stage Setting

3.1 The Various Senses of "Confirmation"

"Confirmation" is a multiply ambiguous term in the present context. In this section of the handout, I want to layout some distinctions and some terminology that will (hopefully) help us keep track of the multiple senses and connotations of the term, as we'll be using it in the seminar. There are three main dimensions along which the term "confirmation" will be used here: *logical*, *epistemic*, and *psychological* (or *cognitive*).

- **Logical**. In its *logical* sense, "confirmation" denotes a relation between *statements*.[2] Specifically, it will be (at least[3]) a three-place relation between *premise* (P), *conclusion* (C), and *background corpus* (K). The basic idea here is (roughly) that deductively valid arguments are just "special cases" of "logically good" arguments. Take the simplest sentential-logical case. An argument in sentential logic from P to C is *valid* (relative to background corpus K) just in case *all* truth-value assignments which assign $\top$ to P (and K) also assign $\top$ to C. Here's a rather naïve way to generalize sentential validity:

  > An argument from P to C is "inductively strong" (relative to K) — *i.e.*, P *confirms* C (relative to K) — iff "most" truth-value assignments which assign $\top$ to P (and K) also assign $\top$ to C.

  As it turns out (for various reasons that we'll discuss later in the seminar), this isn't such a good way to generalize the sentential validity concept. But, it is a *generalization* of validity, since all (but not only) valid arguments come out "strong" or "confirmatory" according to this account. So far, all we have here is a *qualitative* (yes/no) characterization of logical confirmation. We could also imagine comparative and even quantitative concepts. For instance, a naïve comparative account might say:

  > The argument from P_1 to C_1 is "stronger" (relative to K_1) than the argument from P_2 to C_2 is (relative to K_2) iff $v_1 < v_2$, where v_1 is the "proportion" of truth-value assignments which assign $\top$ to P_1 (and K_1) but $\perp$ to C_1, and v_2 is the "proportion" of truth-value assignments which assign $\top$ to P_2 (and K_2) but $\perp$ to C_2.

  We could even take "the 'proportion' of truth-value assignments which assign $\top$ to P (and K) but $\perp$ to C" to be a quantity that is "inversely proportional" to the "strength" of the argument from P to C (relative to K). Of course, I'm not presenting any of these ideas as genuine *accounts* of logical confirmation (although, as we'll see, accounts not too different from these have actually been advocated in the literature). I'm simply giving you a sense of what such a thing might look like.

  Notationally, I will adopt the following conventions for the seminar.

  - $\mathfrak{C}(C, P \mid K) \overset{\text{def}}{=} P$ confirms C, relative to K. [qualitative concept]
  - $\mathfrak{M}(\langle C_1, P_1 \mid K_1\rangle, \langle C_2, P_2 \mid K_2\rangle) \overset{\text{def}}{=} P_1$ confirms C_1, relative to K_1 more strongly than (or, perhaps, "better than") P_2 confirms C_2, relative to K_2. [comparative concept]
  - $\mathfrak{c}(C, P \mid K) \overset{\text{def}}{=}$ the degree to which P confirms C, relative to K. [quantitative concept]

  As we'll see, Carnap introduces, basically, this taxonomy of logical confirmation concepts (with slightly different notation). Carnap also assumes that the quantitative concept ($\mathfrak{c}$) is the most basic, and that we can define the comparative ($\mathfrak{M}$) and qualitative ($\mathfrak{C}$) concepts in terms of it. This reductionistic assumption is controversial (as we'll see). But, the taxonomy is very useful in any event. I will also

[2] I am intentionally being unclear here as to whether I mean *sentences* or *propositions*. We'll come back to that issue later. I am also simplifying here a bit, since I should probably talk about *sets* (or *multisets)* of statements. Again, we'll come back to that.

[3] Later, I will suggest that the logical confirmation relation is a four-place relation. For now, three places will be enough. The important point here is that none of the relata of the logical confirmation relation are *agents* (actual or ideal).

recommend that we extend this taxonomy to the epistemological and psychological confirmation concepts. When we do (below), however, we will add subscripts to indicate that we're no longer talking about the *logical* confirmation concepts. Hopefully, this will give us a less ambiguous way of talking about the various kinds of confirmation floating around in the contemporary literature.

- **Epistemological**. In its *epistemological* sense, "confirmation" denotes a relation that involves an *epistemic agent* ϕ (usually an *idealized* one, which is assumed to be "epistemically rational"). There are various sorts of relations (and other relata) that may arise in the epistemological setting. Examples:
  - Degree of epistemological confirmation as *degree of justified belief*. On this reading, the relata of the relation will — like the logical relation — include three statements (or, perhaps, some sort of cognitive states[4]) *plus an idealized agent*. Specifically, we might have the following epistemological-confirmation-as-degree-of-justified-belief concept.[5]
    * $\mathfrak{c}_\phi(C, P \mid K) \overset{\text{def}}{=}$ the degree to which P epistemologically confirms C, relative to K, *for epistemic agent* $\phi \overset{\text{def}}{=}$ the degree to which ϕ is justified in believing C, on the basis of P, relative to K.
  - Degree of epistemological confirmation as *degree evidential support*. Here, we might have:
    * $\mathfrak{c}_\phi(C, P \mid K) \overset{\text{def}}{=}$ the degree to which P evidentially supports C, relative to K, *for agent* ϕ.

  Following Carnap, for each such quantitative epistemological confirmation concept, we can imagine corresponding qualitative and comparative epistemological confirmation concepts ($\mathfrak{C}_\phi$ and $\mathfrak{M}_\phi$). As we will see, there is a fair amount of confusion in the literature about the distinction between degree of belief and degree of evidential support. We will talk a lot about this later in the course.

- **Psychological (Cognitive)**. In its *psychological* sense, "confirmation" denotes a relation that involves a *cognitive agent* ψ. Unlike the epistemic agent ϕ, the cognitive agent ψ is not assumed to be ideally rational in any (thick) *normative* sense. But, the cognitive agent will be assumed to be "rational" in *some* (broadly, teleological) sense — at least to the extent that their cognitive systems are *functioning properly*. This leaves ample room for significant divergences between ϕ's and ψ's. The extent of the divergence between ϕ's and ψ's is an important issue in the psychological confirmation literature we will be reading. As in the epistemic case, I will assume that the relata of the confirmation relation (other than the agent) are *statements* (and, again, this will be controversial in some contexts). E.g.,
  - Degree of psychological confirmation as *actual degree of belief*.
    * $\mathfrak{c}_\psi(C, P \mid K) \overset{\text{def}}{=}$ the degree to which P psychologically confirms C, relative to K, *for cognitive agent* $\psi \overset{\text{def}}{=}$ the degree to which ψ actually believes C, on the basis of P, relative to K.
  - Degree of psychological confirmation as *actual degree of evidential support*.
    * $\mathfrak{c}_\psi(C, P \mid K) \overset{\text{def}}{=}$ the degree to which P evidentially supports C, relative to K, *for agent* ψ.

  Again, for each such quantitative psychological confirmation concept, we can imagine corresponding qualitative and comparative epistemological confirmation concepts ($\mathfrak{C}_\psi$ and $\mathfrak{M}_\psi$). Moreover, we can imagine *eliciting* such degrees of psychological confirmation from actual agents. In Part II of the course, we'll see experiments that have been designed to elicit (or that have been *interpreted as* "accidentally" eliciting) such degrees of psychological confirmation.

3.2 Some Substantive Theoretical Background

3.2.1 Classical Deductive Logic

I will assume intimate familiarity with sentential logic and monadic predicate logic (we will see a bit of relational predicate logic, but not very much of it). For those of you who are not intimately familiar with these classical logical theories, I would recommend not taking this course.

[4] There is some controversy in epistemology as to what the relata of epistemological relations are (other than an idealized agent). For simplicity, I will assume that the relata are *statements* and an idealized agent. We may have reason to worry about this later.

[5] For simplicity, I am suppressing the *epistemic context* in which the agent ϕ finds herself. We can, for now, just take that to be encoded in a suitable description of ϕ, which is to be taken into account in the determination of $\mathfrak{c}_\phi$.

3.2.2 Probability Calculus

Ideally, you are already familiar with probability calculus. If not, I recommend reading Ellery Eells's "crash course" on probability calculus that I have on the syllabus (for this week). I also recommend looking at my PHIL 148 course website (the first few weeks of it) for extensive materials on basic probability calculus. Meanwhile, I will reproduce some of that material here as well. For simplicity, I will only talk about probability calculus for sentential languages (the extension to the monadic predicate case is straightforward).

One way to think about probabilities is *via* the *axioms* of probability calculus. Assume we have a sentential language $\mathcal{L}$. A probability function on $\mathcal{L}$ is a function $\Pr(\cdot)$ from $\mathcal{L}$ to the real numbers, such that:

1. For all $p \in \mathcal{L}$, $\Pr(p) \geq 0$.
2. For all $p \in \mathcal{L}$, if $p \dashv\vDash \top$, then $\Pr(p) = 1$. [where $\top$ is any tautology of $\mathcal{L}$, and $\dashv\vDash$ is tautological equivalence]
3. For all $p, q \in \mathcal{L}$, if $p \,\&\, q \vDash \bot$, then $\Pr(p \vee q) = \Pr(p) + \Pr(q)$. [$\bot$ is self-contradictory, $\vDash$ is entailment]

It can be shown that this axiomatization is equivalent to the one Eells discusses (which is Kolmogorov's). Given our axioms for $\Pr(\cdot)$, we then *define conditional* probability $\Pr(\cdot \mid \cdot)$ in terms of $\Pr(\cdot)$, as follows:

4. For all $p, q \in \mathcal{L}$, $\Pr(p \mid q) \overset{\text{def}}{=} \dfrac{\Pr(p \,\&\, q)}{\Pr(q)}$, provided that $\Pr(q) > 0$.[6]

These axioms are sufficient to characterize probability calculus (as we will use the term in this course). But, they are rather unwieldy to use. Another way to think about probability calculus is "semantically" (or algebraically). This involves the use of "stochastic truth-tables" and high-school algebra. Here's an example. Consider a sentential language $\mathcal{L}_2$ with 2 atomic sentences "X" and "Y". A *stochastic truth-table* for $\mathcal{L}_2$ is:

X	Y	State Descriptions	$\Pr(s_i) = \mathfrak{s}_i$
$\top$	$\top$	$s_1 \overset{\text{def}}{=} X \,\&\, Y$	$\mathfrak{s}_1$
$\top$	$\bot$	$s_2 \overset{\text{def}}{=} X \,\&\, {\sim}Y$	$\mathfrak{s}_2$
$\bot$	$\top$	$s_3 \overset{\text{def}}{=} {\sim}X \,\&\, Y$	$\mathfrak{s}_3$
$\bot$	$\bot$	$s_4 \overset{\text{def}}{=} {\sim}X \,\&\, {\sim}Y$	$\mathfrak{s}_4$

As long as the numbers ($\mathfrak{s}_i$) assigned to each state description (s_i) are all on $[0, 1]$ and sum to 1, then this will define a probability function $\Pr(\cdot)$ on $\mathcal{L}_2$ [*e.g.*, if each state description is equiprobable, then we will have $\Pr(s_i) = \mathfrak{s}_i = 1/4$, for all i]. That is, once we have the "basic probabilities" $\mathfrak{s}_i$ specified for each of the state descriptions s_i, this will determine a probability for *every* sentence $p \in \mathcal{L}$. This is because every sentence in $\mathcal{L}$ is equivalent to some disjunction of state descriptions. For instance, $X \dashv\vDash s_1 \vee s_2$. Thus, $\Pr(X) = \Pr(s_1 \vee s_2) = \mathfrak{s}_1 + \mathfrak{s}_2$. In general, the probability of $p \in \mathcal{L}$ is the sum of the probabilities of the state descriptions contained in the disjunction of state descriptions that is equivalent to p. This algebraic representation of sentential probability calculus is explained in detail in my paper "A Decision Procedure for Probability Calculus with Applications" (posted on the syllabus, for the Week 1 readings). In that paper, I also show that this algebraic representation of sentential probability calculus facilitates a decision procedure for sentential probability calculus. This means that sentential probability calculus is a decidable theory (just as sentential logic is). One can use this algebraic representation to prove theorems and find counterexamples to non-theorems of probability calculus. In many cases, this algebraic method is easier to understand (and work with) than the axiomatic method. Indeed, I have written a computer program (called PrSAT) that implements an algebraic decision procedure for probability calculus (analogous to a truth-table generator for sentential logic). This is a useful tool for working with probability calculus. I will discuss some applications of PrSAT in confirmation theory, later in the course. If you're interested in using PrSAT, I'd be happy to help you set it up on your computer. Meanwhile, I recommend trying to develop some degree of familiarity with both the axiomatic and algebraic approaches to probability calculus (working through the theorems Eells proves axiomatically, reading my paper on PrSAT, and/or studying my lecture notes for the first few weeks of PHIL 148 should get you off to a very good start). This will help you evaluate some of the technical probabilistic arguments we'll encounter in the literature we'll be reading. While that won't be a huge part of the course (overall), it is a rather important aspect of the course.

[6] Things get tricky (and our definition of $\Pr(p \mid q)$ goes silent) when $\Pr(q) = 0$. We'll talk about that later in the course.