Scientific Explanation & Scientific Realism: Day 1

- Administrative:
  - We meet at last!
  - 2 cancellations, but 14 meetings (including last class day)
  - Perhaps we'll do one make-up (some Wed. or Fri.)
  - Enrolled students see me after class (brief admin. stuff)
  - Website updated with readings through week 8
    `http://socrates.berkeley.edu/~fitelson/290/`
- Brief Introductory/Overview Remarks About the Seminar
- Hempelian Theory: The Deductive-Nomological (D-N) and Inductive–Statistical (I–S) Accounts of Scientific Explanation
  - H & O's formal D-N Account & problems
  - The Explanation/Prediction *Symmetry Thesis* & problems
  - Next: The (I–S) Account, the (S–R) account & problems

Deductive–Nomological (D–N) Account of Scientific Explanation I

- Hempel & Oppenheim laid the foundation for contemporary analytic philosophical thought about scientific explanation. Their D–N model is "the fountainhead." H & O start with four D–N adequacy conditions:

DN1. A D–N explanation must be a deductively valid argument.

DN2. The *explanans* must contain – *essentially* (to the entailment of the *explanandum*) – at least one general, *lawlike* sentence.

DN3. The *explanans* must have empirical content (*i.e.*, the *explanans* must be empirically testable or confirmable, in principle).

DN4. The sentences constituting the *explanans* must be true.

- Note: These conditions allow for the case in which a less general "law" (Kepler's) is explained (subsumed) by more general laws (Newton's).
- In order to be clear on what these conditions of adequacy require, we must say more about what general, "lawlike sentences" are . . .

The D–N Account of Scientific Explanation II

- H & O give some guidance on "lawlike sentences":

L1. Lawlike sentences have *universal* ($\forall$) form.

L2. Their scope is *unlimited*.

L3. They do *not* contain designations of *particular* objects.

L4. They contain *only* purely *qualitative* predicates.

- (L1) and (L2) require laws of nature to be *universal*, and to range over the *entire universe*. Why not allow $\exists$'s? Couldn't there be $\exists$-laws?
  - Newton's laws are $\forall$, and they range over all objects in the universe.

(*) All the quarters in John's pocket are made of silver.

- We do *not* want to call sentences like (*) laws of nature. This is *partly* because (*) makes reference to *particular* objects in the (actual) world.
- Sentences can also make *implicit* reference to (actual) particulars, by using *non-qualitative* predicates like "planetary", "arctic", "American".

The D–N Account of Scientific Explanation III

- H & O's (L1)–(L4) concerning laws seem to be *both* too weak *and* too strong. They seem *too weak* because they do not include *modality*.
- Laws of nature seem to have *modal force*. They tell us not only what *happens to be* true in the actual world, but what *must* be true — in *all physically, or nomologically possible worlds*. Consider the following:

(i) No signal travels faster than the speed of light.
(ii) No gold sphere has a mass greater than 100,000 Kg.
(iii) No uranium sphere has a mass > 100,000 Kg.

- Sentences (i) and (iii) seem lawlike. But, does (ii)? Sentence (ii) may *happen to be* true in the actual world. But, sentences (i) and (iii) are *nomologically necessary* — they're true in *all physically possible worlds*.
- Lawlike sentences seem to support *counterfactuals*. (*) does *not* support the counterfactual "if this (non-silver) quarter *were* in John's pocket, then it *would* be made of silver". Why not? Do Newton's laws?

The D–N Account of Scientific Explanation IV

At this point, we need some linguistic terminology (from first-order logic):

- An *atomic sentence* is one that contains no quantifiers, no variables, and no logical connectives (*e.g.*, "Ra", "Lbc", or "$Bdef$").
- A *basic sentence* (also called a "literal") is either an atomic sentence or the negation of an atomic sentence (*e.g.*, "Ra", "$\sim Rb$", *etc.*).
- *Singular sentences* are just *molecules* formed out of basic sentences and logical connectives (*e.g.*, "$Ra \,\&\, Ba$", or "$Lcd \vee \sim Rghi$").
- A *generalized sentence* contains one or more quantifiers followed by an expression containing no quantifiers (*e.g.*, $(\forall x)(\exists y)\ Lxy$).
- A *universal sentence* is generalized using *only* universal quantifiers ($\forall$).
- A sentence is *purely* generalized/universal if it uses no proper names.
- A sentence is *essentially* generalized/universal if it is generalized / universal, *and* it is not equivalent to any singular sentence.

The D–N Account of Scientific Explanation V

- H & O's (L1)–(L4) are *too strong* because they rule-out (so-called) "phenomenological laws" like Kepler's laws of planetary motion.
- H & O are aware of this. For this reason, they make a distinction between "derivative laws" and "fundamental laws".
  - A *derivative law* is a sentence that is essentially, but not purely, universal and is deducible from some set of fundamental laws.
  - A *law* is any sentence that is either fundamental or derivative.
- Kepler's laws of planetary motion are *derivative* laws. They are *not fundamental* laws, because they implicitly use *proper names* (*i.e.*, "Mars", "Earth", *etc.*). Newton's laws are *fundamental* (*i.e.*, essentially *and* purely generalized), and from them we can derive Kepler's laws.
- We can give a D–N explanation in which Newton's laws are among the explanans, and Kepler's are the explanandum. In this sense, the D–N model can undergird our intuition that Newton's laws *explain* Kepler's.

The D–N Account of Scientific Explanation VI

- In the official, formal statement of their theory of explanation, H & O do not use the concept of a law at all. Instead, they move to talk of *theories*. Can you see the difference? Hint: generalized *vs.* universal.
  - A *fundamental theory* is any purely generalized and true sentence.
  - A *derivative theory* is any sentence that is essentially, but not purely, generalized and is derivable from fundamental theories.
  - A *theory* is any fundamental or derivative theory.
- According to these definitions, every law is a theory (but *not conversely*), and every theory is true. Why make every theory true?
- The difference between laws and theories is that theories may contain existential quantifiers ($\exists$), but laws may not (laws must be *universal*).
- H & O require all explanatory theories to be *general* (but *not necessarily universal*) and *true*. As we'll see, these assumptions have (by and large) remained in the contemporary literature on explanation.

The D–N Account of Scientific Explanation VII

- Now, we're ready for the official, formal statement of the D–N theory of scientific explanation (in a few stages):
  - $\langle T, C\rangle$ is a potential explanans of E (a singular sentence) *only if*
    1. T is essentially general and C is singular, and
    2. E is derivable from T and C jointly, but not from C alone.
  - Note: this is *only* a *necessary* condition for $\langle T, C\rangle$'s being a potential explanans of E. If it were taken to be *sufficient*, then we would have any E explained by any true lawlike statement T!
  - Let E be "Mount Everest is snowcapped", T be "All metals are good conductors of heat", and C be "$T \rightarrow E$". Then, both (1) and (2) are satisfied, and so $\langle T, C\rangle$ would be a potential explanans of E.
  - This is absurd, since we have a fact about Mount Everest being explained by a law concerning the heat conductivity of metals!
  - We need to add a further constraint to our definition...

The D–N Account of Scientific Explanation VIII

- H & O add the following condition, to block this triviality:
  3. T must be compatible with at least one class of basic sentences which has C but not E as a consequence.
- In other words, (3) says that for any given theory T, there must be a way to verify that C is true without also *automatically* verifying (confirming) that E is true as well. This yields the following *definition:*
  - $\langle T, C\rangle$ is a potential explanans of E (a singular sentence) *iff*
    1. T is essentially general and C is singular, and
    2. E is derivable from T and C jointly, but not from C alone.
    3. T must be compatible with at least one class of basic sentences which has C but not E as a consequence.
- It is a small step from this definition of a "potential explanans" to the official (complete) definition of a D-N explanation ...

The D–N Account of Scientific Explanation IX

- Finally, here's the official definition of a D–N explanation:
  - $\langle T, C\rangle$ is an explanans of E (a singular sentence) *iff*
    1. $\langle T, C\rangle$ is a potential explanans of E
    2. T is a theory, and C is true.
- Taken together, the explanans $\langle T, C\rangle$ and the explanandum E constitute a D–N explanation of E. This completes the Hempel & Oppenheim explication of the *D–N explanation of a particular fact.*
- Surprisingly, even this careful rendition of D–N explanation is not quite technically correct. Kaplan, Montague, and others give the following counterexample, which they claim shows that $\langle T, C\rangle$ can D–N explain E (on the above account) even if $\langle T, C\rangle$ is *utterly irrelevant to* E.
- Let T be "$(\forall x)Ix$" (*e.g.*, everyone is imperfect), and let E be "Ma" (*e.g.*, Hempel is male). Intuitively, T is completely irrelevant to E.

The D–N Account of Scientific Explanation X

- Next, we deduce the following (derivative) theory from T: (T') $(\forall x)(\forall y)[Ix \vee (Py \rightarrow My)]$. Is it obvious that T' is irrelevant to E?
- And, we choose as our singular sentence: (C) $(Ib \vee {\sim}Pa) \rightarrow Ma$.
- To keep things concrete, let b denote Oppenheim, and let "Px" mean "x is a philosopher". Now, it can be shown (Salmon does so, and I'll sketch it next) that $\langle T', C\rangle$ is an explanans of E — in the D–N sense defined above (*i.e.,* (1)–(3), plus T' is a theory and C is true).
- This is considered to be a "counterexample" to the D–N account. **Exercise**: try to explain why (or argue the contrary!) this is "bad news" for Hempel and Oppenheim's D–N theory of explanation.
- Jaegwon Kim (1963) suggests adding a fourth condition to "fix" this:
  4. E must not entail any conjunct in the conjunctive normal form of C.

  Exercise: show (4) *is* a fix, and discuss the *consequences* of (4).

Kaplan *et al*'s "Counterexample" I

- To establish that $\langle T', C\rangle$ is a D–N explanation of E, we must show that:
  1. T' is true and essentially general and C is true and singular. T' says "every pair of people $\langle x, y\rangle$ is such that either x is imperfect or y is a non-philosopher or y is male." This is general and true [since $(\forall x)\ Ix$]. C says "Oppenheim is perfect or Hempel is a philosopher or Hempel is male." This is singular and true (since Ma).
  2. $T' \,\&\, C \vDash E$, but $C \nvDash E$. Salmon shows $T' \,\&\, C \vdash E$ in a natural deduction system for first order logic. Can you give a simpler proof? It is clear that $C \nvDash E$, since "Oppenheim is perfect" $\vDash C$, but $\nvDash E$.
  3. T' is compatible (*i.e.*, logically consistent) with at least one class of basic sentences which has C but not E as a consequence. To see this, choose (as Salmon does) the set of basic sentences: $\{{\sim}Ib, Pa\}$. Can you see why T' is consistent with ${\sim}Ib \,\&\, Pa$? And, can you see why ${\sim}Ib \,\&\, Pa \vDash C$, but ${\sim}Ib \,\&\, Pa \nvDash E$? See Salmon for the former.

Kaplan *et al*'s "Counterexample" II

T'. Every pair of people $\langle x, y\rangle$ is such that either x is imperfect or y is a non-philosopher or y is male.

C. The pair ⟨Oppenheim, Hempel⟩ is such that either Oppenheim is perfect or Hempel is a philosopher or Hempel is male.

E. $\therefore$ Hempel is male.

- Why do we have doubts about thinking of this as an *explanation* of E?
- Kim suggests a "fix" — add the following condition to D–N definition:
  4. E must not entail any conjunct in the conjunctive normal form of C.
- In this case, the conjunctive normal form (CNF) of C is:

$$(\sim Ib \vee Ma)\,\&\,(Pa \vee Ma)$$

- So, E entails *both* conjuncts of the CNF of C. What are the *consequences* of adding Kim's (4)? Does it rule-out *too many* potential explanations? Is there independent motivation for (4), or can it only be motivated as a counterexample-blocker? Why not add $E \not\models C$ instead?

The D–N Account of Scientific Explanation XI

- Things needed to complete the D–N Account:
  1. Explications of model(s) of *probabilistic* or *statistical* explanation
  2. An adequate (D–N) account of the explanation of *laws*. On the current account, a derivative law L can be "explained" by the conjunction $L\,\&\,L'$, for *any* L', *no matter how irrelevant to* L' *may be to* L. ["fn. 33"]
  3. A good explication of the concept of a *qualitative* predicate ("grue"?).
  4. A good explication of the concept of a law of nature.
- Potential problems within the underlying D–N framework:
  1. Are (*all*) explanations *arguments*, as H & O assume?
  2. Must all explanations make essential use of *law*(s) of nature?
  3. According to H & O, all (D–N) explanations are (potential) (H–D) predictions, and *vice versa*. Is this *symmetry thesis* correct?
  4. According to H & O, *causality* plays no essential role in the scientific explanation of particular, token events. Is this correct?
  5. Must the explanans of a good explanation be (*literally*) true?

Hempel's Explanation/Prediction Symmetry Thesis

- Hempel thought that every adequate (singular) explanation is a potential prediction, and *vice versa*. Why did he think this? Does it make sense? There are two distinct directions to this thesis:

(⇒) Adequate Explanation ⇒ Potential Prediction

(⇐) Potential Explanation ⇐ Adequate Prediction

- Both directions may sound strange, at first. Some clarification helps:
- Hempel: an adequate explanation should provide grounds for holding that the explanandum was to be expected. So, if E is not known, an explanation should provide grounds that E is to be expected.
- Hempel: (⇒) is "almost trivial." Defense against Scriven's attack:
  - Scriven: Only people who have had syphilis can contract paresis. But, only a small fraction (around 25%) of syphilis patients contract paresis. It seems quite *explanatory* to say that a person got paresis *because* they had syphilis. But, does syphilis *predict* paresis?

  - Hempel: "Precisely because paresis is such a rare sequel of syphilis, prior syphilitic infection cannot by itself provide an adequate explanation for it." This presupposes there are other factors that determine which untreated cases of syphilis go on to develop paresis.
  - But what about genuine chance events – can they can not be explained at all? *E.g.*, the radioactive decay of a particle (at a time). This may be very improbable, given our best theory. But, do we then conclude that our best theory cannot explain these events?
- As for (⇐), Hempel gives what appears to be a counterexample:
  - "A finite set of data obtained in an extensive test of the hypothesis that the electric resistance of metals increases with their temperature may afford good support for that hypothesis and may thus provide an acceptable basis for the prediction that in an as yet unexamined instance, a rise in temperature in a metal conductor will be accompanied by an increase in resistance. But if this event actually occurs, the test data [or the correlation or law one might infer from the data] do not provide an explanation for it."

  - Common Causes: we can *adequately predict* a storm using a falling barometer needle, but is this a potential *explanation* of its arrival?
  - Hempel: there may be incompletely formulated potential explanations in these cases. He doesn't hint at what they might be.

- So, why does Hempel find the symmetry thesis attractive? Consider:

  (‡) $\langle T, C\rangle$ *explains* $E \Leftrightarrow E$ *confirms* $\langle T, C\rangle$ [say E confirms C, given T]

- Hempel's theory of confirmation is also deductive (much like D–N). This is why Hempel feels the need to defend something like (‡).
- It seems that $\Leftarrow$ of (‡) is false, since "explains" seems asymmetric, while "confirms" (*i.e.*, "provides evidence for") is (arguably) symmetric. We'll return to the symmetry question repeatedly this term.
- But, what about $\Rightarrow$ of (‡)? It's not clear that $\Rightarrow$ is false. Take the paresis (Pa) and syphilis (Sa) example. Let T be that "law" that S is (generally) nomically necessary (but only "25% sufficient") for P.
- Then, $\Pr(Pa \,|\, Sa \,\&\, T) = 0.25$, which is low. But, $\Pr(Pa \,|\, {\sim}Sa \,\&\, T) = 0$, which is *much lower*! So, plausibly, Pa *confirms* Sa, given T.

Some Famous Problematic Cases for D–N (& some other theories)

1. **The Eclipse**: One can D–N-explain a current total eclipse, using (say) Newton's laws of motion, together with past positions of the earth, sun, moon. But, one can also D–N-explain a current eclipse by appeal to NL plus *future* positions! Should this count as an *explanation*? Many laws are time-symmetric – perhaps some explanations aren't.
2. **The Flagpole**: We may D–N-explain the length of a shadow cast by a flagpole using certain laws of optics/geometry, together with the position of the sun in the sky, etc. But, we can also D–N-explain the height of the flagpole using the same laws, together with the length of the shadow and the position of the sun! Is this an *explanation*?
3. **The Barometer**: A falling barometer (together with the appropriate meteorological laws) can reliably predict an approaching cold front. So, one may also be able to D–N-explain the approach of the cold front by appealing to the barometer's drop, together with these same meteorological laws. These are *common cause* (*conjunctive fork*) cases.

4. **The Moon and the Tides**: The (general and lawlike) *correlation* between the moon's position and the tides was well known for centuries before Newton's gravitational theory was known. So, reliable predictions, and D–N-explanations of the tides were constructible by these ancestors of Newton. But, arguably, until the *causal story* behind the tides was told, no *legitimate* explanation was really available.
5. **Syphilis and Paresis**: Only people who have had syphilis can contract paresis. But, only a small fraction (around 25%) of syphilis patients contract paresis. It seems quite *explanatory* to say that a person got paresis *because* they had syphilis. But, this cannot be said on a D–N account (which requires *deduction* of each token case). Similar examples arise surrounding quantum-mechanical phenomena.
6. **The Hexed Salt**: Why did this sample of table salt dissolve in this cup of water? Because a person wearing a funny hat mumbled some non-sense syllables and waived a wand over it. That is, the table salt dissolved because it was hexed. And, it is a *law* that all hexed table salt dissolves when placed in water. This fits the D–N pattern ...

7. **Fred Fox on the Pill**: Fred Fox (a male) has not become pregnant during the past year because he has faithfully consumed his wife's birth control pills. And, any male who faithfully takes birth-control pills will avoid becoming pregnant. But, intuitively, there seems to be no *explanatory relevance* in this fact. This also fits the D–N pattern.
8. **Explanation by false/idealized theories**: We use Newton's theory all the time to explain various phenomena. But, we know Newton's theory is *false*. Moreover, for all we know, all of our current scientific theories are also false (in some subtle and as yet unseen way). Would this imply none of our current scientific explanations are good ones?
9. **Logical Tricks & *Misc.* (some like Kaplan *et al*)**:

All things born die.
John was born.
∴ John died.

All crows are black.
John isn't black.
∴ John isn't a crow.

All copper conducts electricity.
Either my alarm didn't go off today, or some copper doesn't conduct electricity.
∴ My alarm didn't go off today.