Four Decades of Scientific Explanation

- Administrative:
  - Try to turn in first drafts soon ...
  - I will return comments within a week of receipt.
  - My office hours today are canceled (another Friedman talk!).
- Some Introductory Remarks
  - Terminology: Explanandum, explanans, etc.
  - Prediction (confirmation) versus explanation
- The Deductive-Nomological (D-N) Account of Scientific Explanation — The First "Serious" Theory of Explanation
  - Hempel & Oppenheim's Conditions of Adequacy
  - What are (nomological) "lawlike sentences"?
  - H & O's formal D-N Account, and some of its problems

Explanation *versus* Prediction (Confirmation) I

- Prediction (confirmation) involves providing reasons to believe that (or evidence that) certain claims about observables (theories) are true.
- Explanation involves answering questions of the form "Why X (in K)?" (or "How X (in K)?"), where "X" is *assumed to be true*.
- The *explanandum* of an explanation is that which is being explained, and the *explanans* of an explanation is that which does the explaining.
- That is, the explanandum is the "X" in "Why X?", and the explanans is the (an) *answer* to the explanation-seeking why question.
- The explanandum is *assumed to be true* (in the context). And, so the explanans need not give reason to believe that X is the case.
- We're not looking to be *convinced* that X is true, we just want to know *why* (how) X is true. Therefore, intuitively, "p explains q" is different from "p predicts (or confirms/is confirmed by) q".

Explanation *versus* Prediction (Confirmation) II

- Here are some intuitive examples which should illustrate the differences between prediction/confirmation and explanation:
  - A falling barometer may *confirm* an approaching cold front, but it does *not explain why* the cold front is approaching.
  - The length of a shadow (cast by a flagpole of a certain height) may *confirm* the sun's position in the sky, but it does *not explain* it.
  - The *anthropic principle* says that we my safely *infer* (*i.e.*, we may *predict* / retrodict) certain things about the history of our universe from the fact that we now exist (*e.g.*, we know that certain conditions favorable to the existence of life in the universe must have evolved). But, the anthropic principle does *not* say that our present existence *explains why* our universe evolved the way it did.
  - Can you think of a similar example to illustrate the distinction?

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

- Hempel *&* Oppenheim (1948) laid the foundation for contemporary analytic philosophical thought about scientific explanation. Their D–N model is "the fountainhead." H *&* O start with 4 adequacy conditions:
  1. A scientific explanation must be a *deductive*ly valid argument.
  2. The explanans must contain — *essentially* — at least one (*nomological*) general *lawlike* sentence.
  3. The explanans must have empirical content (contrast with "pure mathematical explanation" — see Mark Steiner's paper).
  4. 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 (nomological) "lawlike sentences" are ...

The D–N Account of Scientific Explanation II

- H *&* O give some guidance on (nomological) "lawlike sentences":
  1. Lawlike sentences have *universal* ($\forall$) form.
  2. Their scope is *unlimited*.
  3. They do *not* contain designations of *particular* objects.
  4. They contain *only* purely *qualitative* predicates.
- (1) and (2) 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 "lunar", "arctic", or "American".

The D–N Account of Scientific Explanation III

- H *&* O's (1)–(4) seem to be *both* too weak *and* too strong. They are *too weak* because they do not include *modality*.
- Laws of nature 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*.
  - (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) are lawlike. But, (ii) is not. 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 must 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 (1)–(4) 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 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 very careful rendition of D–N explanation is not quite technically correct. Kaplan, Montague, and others (1961) give the following counterexample, which 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 "$(\forall x)Fx$" (*e.g.*, everyone is imperfect), and let E be "Ha" (*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 theory from T: (T') "$(\forall x)(\forall y)[Fx \vee (Gy \to Hy)]$". Is it obvious that T' is irrelevant to E?
- And, we choose as our singular sentence: (C) "$(Fb \vee {\sim}Ga) \to Ha$".
- To keep things concrete, let b denote Oppenheim, and let "Gx" mean "x is a philosopher". Now, it can be shown (although, I won't do so here — see Salmon) 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. **Paper topic**: 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.

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

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.
  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?