Scientific Explanation — Background Notes

- Administrative:
  - Introductions...
  - Syllabus, course website, etc...
- Some Introductory Remarks
  - Terminology: Explanandum, explanans, etc.
  - Prediction (confirmation) versus explanation
- Some of the historical background leading up to Woodward's book
  - Hempel & Oppenheim's Fountainhead D–N Account (and problems)
  - Statistical/Probabilistic Descendants of H & O (and problems)
  - I have posted a bunch of salient background readings on the website
  - These background notes are, basically, the first 2–3 "decades" in Salmon's *Four Decades of Scientific Explanation* (posted online)
  - I won't get into the "causal" approaches to scientific explanation in these notes. For that, we'll just jump right into Woodward's book.

Explanation *versus* Prediction (Confirmation) I

- *Prediction* (*confirmation*) involves providing reasons to believe that (or evidence that) certain claims (specifically, *scientific theories*) are true.
- *Explanation* involves answering questions like "Why (or how) is it the case that X?", where "X" is *assumed to be true* (in the context C in which the question is asked). The context C also sets-up *contrasts* and *background conditions* that are explanatorily relevant/salient (see Bromberger and van Fraassen for more on contrastivity).
- 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 some explanation-seeking why question.
- The explanandum is *assumed to be true* (in the context C). And, so the explanans need not give reason to believe that X is the case.

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.
  - Many other examples can be used to illustrate this 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" — which we will not be discussing).
  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* (∀) form.
  2. Their scope is *unlimited*.
  3. They do *not* contain designations of *particular* objects.
  4. They contain *only* "purely *qualitative*" predicates (natural kinds?).
- (1) and (2) require laws of nature to be *universal*, and to range over the *entire universe*. Why not allow ∃'s? Couldn't there be ∃-laws?
  - Newton's laws are ∀, and they range over all objects in the universe.

  (*) All the quarters in John's pocket are made of silver.
- Some would *not* want to call (*) a *law* 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) may seem to be *both* too weak *and* too strong. They may seem to be *too weak* because they do not take *modality* into account.
- Laws of nature seem to have *modal force*. They seem to tell us not only what *happens to be* true in the actual world, but what *must* be true — in *all (or most) 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.
- Claims (i) and (iii) have modal force. But, (ii) does 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 also *support counterfactuals*. (*) does *not* support (*e.g.*) the counterfactual "if this (non-silver) quarter *were* in John's pocket, then it *would* be made of silver". And (iii)? [Newton's laws?]

The D–N Account of Scientific Explanation IV

For H & O's formal account, we need to introduce some FOL terminology:

- 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 (∀).
- 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) may seem *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* may be either a fundamental law or a derivative law.
- 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 (∃), 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* (singular) E explained by *any* true lawlike statement T!
  - Let $E \stackrel{\text{def}}{=} Fa$, $T \stackrel{\text{def}}{=} (\forall x)(Gx \supset Hx)$, $T' \stackrel{\text{def}}{=} Gb \supset Hb$, and $C \stackrel{\text{def}}{=} T' \supset E$. Then, both (1) and (2) will be satisfied by $\langle T', C\rangle$, provided T is true. Thus, $\langle T', C\rangle$ would be a potential explanans of E (*if* 1&2 sufficed).
  - This is absurd, since we would then have arbitrary singular facts being explained by arbitrary "laws" relating arbitrary properties.
  - 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 explication:
  - $\langle T, C\rangle$ 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 explication of a "potential explanans" to the official (complete) explication of a D-N explanation …

The D–N Account of Scientific Explanation IX

- Finally, here's the official explication 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 a *D–N explanation of a particular fact*.
- Unfortunately, even this very careful rendition of D–N explanation suffers from some bothersome technical difficulties.
- Kaplan, Montague, and others (1961) describe a simple class of cases, which seem to show that $\langle T, C\rangle$ can D–N explain E (on the above account) even if $\langle T, C\rangle$ is (intuitively) *utterly irrelevant to* E.
- Next, I'll discuss an early "counterexample" to D–N, which begins to raise the (vexing and recurring) issue of *explanatory relevance*.

The D–N Account of Scientific Explanation X

- Alleged counterexample to D–N (Kaplan *et al. circa* 1961):
  - Let $T \stackrel{\text{def}}{=} (\forall x)Fx$ (*e.g.*, everyone is imperfect), and let $E \stackrel{\text{def}}{=} Ha$ (*e.g.*, Hempel is male). T is (intuitively) *explanatorily irrelevant* to E.
  - $T \vDash T' \stackrel{\text{def}}{=} (\forall x)(\forall y)[Fx \vee (Gy \supset Hy)]$. Let $C \stackrel{\text{def}}{=} (Fb \vee {\sim}Ga) \supset Ha$, where the constant b denotes Oppenheim (Hempel's co-author).
  - Kaplan, *et al* show that $\langle T', C\rangle$ is a D–N explanation of E. They think this is counter-intuitive…Do you agree with this claim?
  - Next, I will say a bit more about this alleged counterexample, and one proposed way to avoid it (due to Jaegwon Kim).
- 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.
  2. $T' \,\&\, C \vDash E$, but $C \nvDash E$.
  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.

The D–N Account of Scientific Explanation XI

- 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 $(\forall x)\, Fx$ is true]. C says "Either (Oppenheim is perfect and Hempel is a philosopher) or Hempel is male." This is singular and true (since Ha).
  2. $T' \,\&\, C \vDash E$, but $C \nvDash E$.
     - Salmon shows that $T' \,\&\, C \vdash E$ using a natural deduction system for first order logic. Can you give a simpler proof/argument? It is clear that $C \nvDash E$, because "Oppenheim is perfect" $\vDash C$, but "Oppenheim is perfect" $\nvDash E$.
  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: $\{{\sim}Fb, Ga\}$. Can you see why T' is consistent with ${\sim}Fb \,\&\, Ga$? And, can you see why ${\sim}Fb \,\&\, Ga \vDash C$, but ${\sim}Fb \,\&\, Ga \nvDash E$? See Salmon for the former.

The D–N Account XII: "Counterexample" #1 Cont'd

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 and 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?
- Jaegwon Kim suggested a "fix" — add the following *fourth condition*:
  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:

$$({\sim}Fb \vee Ha) \,\&\, (Ga \vee Ha)$$

- So, E entails *both* conjuncts of the CNF of C. Thus, in fact, $E \vDash C$.
- So, why does Kim add a requirement *stronger* than just $E \nvDash C$? Answer: slightly more complicated examples (also intuitive "irrelevancies") exist in which $E \nvDash C$, but E does entail some conjunct in the CNF of C.

Overview of the Kaplanian Relevance Problem for D–N Explanation

- These Kaplanian counterexamples to D–N trade (in part) on the fact that *material* implication does not require the antecedent to be *relevant* to the consequent. Since $P \supset Q \dashv\vDash {\sim}P \vee Q$, we have (infamously):

$$Q \vDash P \supset Q \qquad \text{and} \qquad {\sim}P \vDash P \supset Q$$

- Other conditionals do not imply these "paradoxes" of $\supset$. For instance, *counterfactual* conditionals $P \mathrel{\square\!\!\rightarrow} Q$ require some sort of (nomic) *relevance* of P to Q (*i.e.*, some sort of *nomic dependence* of Q on P).
- Let's apply this to a simple example involving a "covering law". Let $Rx \stackrel{\text{def}}{=} x$ is in this Room, $Ex \stackrel{\text{def}}{=} x$ speaks English, $a \stackrel{\text{def}}{=}$ Branden. Contrast:

$T.\ (\forall x)(Rx \supset Ex)$	$T'.\ (\forall x)(Rx \mathrel{\square\!\!\rightarrow} Ex)$
$C.\ Ra$	$C.\ Ra$
$\therefore E.\ Ea$	$\therefore E.\ Ea$

- Dilemma: (T) is true, but *explanatorily irrelevant* to (E); (T') is (more) explanatorily relevant to (E), but *false*. It's easy (hard) to make $\supset$'s ($\mathrel{\square\!\!\rightarrow}$'s) *true*, but hard (easy) to make $\supset$'s ($\mathrel{\square\!\!\rightarrow}$'s) *explanatorily relevant*.

Some Issues Raised by The D–N Account of Explanation

- Things needed to complete the D–N Account:
  1. An explication of the concept of a *law* (of nature).
  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. An explication of the concept of a "*qualitative* predicate" ("natural kind"?).
  4. Explications of *probabilistic* and/or *statistical* explanation.
- 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 *explanation-prediction symmetry thesis* correct?
  4. Must the explanans of a good explanation be (*literally*) true?
  5. According to H & O, *causality* plays no essential role explanation. More generally, H & O seem to lack a (thick) notion of *explanatory relevance*.

Nine Famous Problematic Cases for D–N (& 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*?
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 H–D-predict an approaching cold front. Thus, one could also D–N-explain the approach of the cold front using the barometer's falling, together with these same meteorological laws.

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, H–D-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. This also fits the D–N pattern.
8. **Joint effects of a common cause**: Consider two TVs receiving a common broadcast (with one TV farther from the source). We can use the closer TV to *predict* what will be seen on the farther TV. But, does this *explain why* this is seen on the farther TV? What does this example suggest about the explanation/prediction *symmetry thesis*?
9. **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). Does this mean none of our current scientific explanations are good ones?

Think about the two directions of the explanation/prediction symmetry thesis [(potential) explanation ⇔ (potential) prediction]. Is either correct? How does this depend on choice of theories of explanation/prediction?

Decade #2 — The Birth of Statistical Explanation

- H & O were aware that their D–N account (as originally stated) left no room for the explanation of either (1) statistical laws, or (2) token events which cannot be derived from any theory (but on which some theory + auxiliaries/initial conditions may *confer a probability*).
- Hempel's first alteration (a very minor one) was to expand the explanation of *laws* (by more general laws) in D–N to the case of *statistical laws*. This led to the *Deductive-Statistical* (D–S) model.
- On the D–S model, a statistical law may be explained by appeal to more general laws (which may be either statistical or universal).
- **Example**: we may derive the half-life of uranium-238 from the basic laws of quantum mechanics (together with the height of the potential barrier surrounding the nucleus and the kinetic energies of the alpha particles within the nucleus). [D–S is a *mere variant* of D–N.]
- Same problems as D–N (*plus statistical* laws & "qualitative predicates"!)