

Session 7. Getting ready to infer: Theoretical Virtues

Dr. Neil Barton

n.a.barton@ifikk.uio.no

FIL4200: Epistemology and Philosophy of Science

13. February 2023

Thus far we've looked at ignorance and suspension of judgement.

We're now going to move to looking at how we might select theories (before we move on to questions of pluralism).

1 Setup

Much of the discussion of theoretical virtues occurs in the context of **realism** vs. **anti-realism**.

Realism. Scientific realism is the view that we ought to believe in the unobservable entities posited by our most successful scientific theories (i.e. our scientific theories are at least **roughly true**).

(e.g. Quine's criterion.)

Anti-realism. Can be more of a cluster of views that reject realism:

- (1.) Reject the metaphysical commitment to the existence of a mind-independent reality
- (2.) Reject the semantic commitment to interpret theories literally or at face value
- (3.) Reject the epistemological commitment to regard theories as furnishing knowledge of both observables and unobservables.

Structural realism. We should commit ourselves to the **mathematical** or **structural** content of our theories (**Note:** This is, as I've set it up, an **anti-realist** position.)

One issue for each view is to examine **theory choice**.

This is often done on the basis of **theoretical virtues**.

Theory choice divides into at least **two** kinds.

Theory choice for application. We want to pick the scientific theory that is most useful for achieving some practical ends.

Theory choice for truth. Pick the (a?) **true** theory.

Clear differences for anti-realists and realists here.

No-miracles argument. Putnam: Realism is the only philosophy that doesn't make the success of science a miracle.

There is a kind of **Inference to the Best Explanation** (IBE) here.

(We will discuss this more next session, so I won't get too into it now.)

On the anti-realist side there's the (among many others):

Pessimistic meta-induction. All our previous theories turned out to be wrong, so our current ones should be too.

Undertermination of theory by evidence. There are many theories that could correspond to the evidence (in particular we can modify the auxiliary assumptions).

Problem of Unconceived Alternatives. Perhaps there are very many good theories that are radically different from our own (this has historical precedent).

If we can get some IBE going, perhaps there's a way of getting some traction on these problems for the realist?

Equally though the anti-realist has clear reasons to be interested in theoretical virtues too.

Questions about the setup?

2 Evidential virtues

Evidential accuracy: A theory T fits the empirical evidence well (regardless of causal claims).

Note: There are some very bad evidentially accurate theories.

Note: There is a difference here between **prediction**, **intended accommodation**, and **unintended accommodation**.

Causal adequacy: T's causal factors plausibly produce the effects (evidence) in need of explanation

Explanatory depth: T excels in causal history depth or in other depth measures such as the range of counterfactual questions that its law-like generalizations answer regarding the item being explained.

Question. How tractable is evidential accuracy?

Question. How tractable are explanatory depth and causal adequacy?

3 Coherential virtues

Internal consistency: T's components are not contradictory.

e.g. Bohr-model of the atom.

An example from mathematics:

The early analysts' noticed that the slope of a line between two points depended upon the *difference quotient*.

For example, if the function we're considering is $f(x) = x^2$, and we take two points x and $x + i$, the difference quotient is:

$$(*) \frac{(x + i)^2 - x^2}{i}$$

We want the limit of this as i approaches 0.

The early analysts' purported solution was as follows: Treat i as a quantity smaller than any finite quantity, but greater than 0.

We can then reason as follows:

(1.) By (*) we have that $f'(x)$ is $\frac{x^2+2xi+i^2-x^2}{i}$ as i approaches 0.

(2.) Simplifying, we get $\frac{2xi+i^2}{i}$.

(3.) Dividing through by i yields $2x + i$.

(4.) But now note that i is smaller than any finite quantity when it approaches 0, so we can eliminate it, obtaining the result that the derivative at x of $f(x)$ is $2x$, as desired.

Who can spot the problem?

Internal coherence: T's components are coordinated into an intuitively plausible whole; T lacks ad hoc hypotheses-theoretical components merely tacked on to solve isolated problems.

e.g. Difficult to measure—e.g. quantum mechanics, what is meant by "intuitively plausible whole"?

Universal coherence: T sits well with (or is not obviously contrary to) other warranted belief.

(e.g. conflict with other theories)

Direct. e.g. Hooke's Law in beam mechanics, directly conflicts when beam too small (also evidentially inaccurate there).

Indirect. e.g. dimension of wave function in QM vs. 4D spacetime in relativistic physics, continuity vs. quanta. Challenge is to integrate two frameworks rather than direct propositional inconsistency.

Closely related: Internal vs. external consistency/coherence (inconsistencies/incoherence within a theory T_1 vs. between two theories T_1 and T_2).

Question. Does this suggest pluralism?

Question. *Should* we be looking to eliminate these incompatibilities?

Question. If we *do* what should the take-away be? (What about when some other cost is paid?)

Question. How strong is the requirement for consistency?

Question. Does this vary with respect to internal vs. external?

Question. What is the *background logic* in which consistency is assessed? (e.g. $P \wedge (Q \vee R) = (P \wedge Q) \vee (P \wedge R)$ is classically valide but invalid in quantum logic).

Question. What about *non-triviality* instead? (e.g. paraconsistent logic).

4 Diachronic virtues

Durability: T has survived testing by successful prediction or plausible accommodation of new data.

Fruitfulness: T has generated additional discovery by means such as successful novel pre-

diction, unification, and non ad hoc theoretical elaboration.

(**Note:** Again the distinction between prediction and accommodation is important here.)

Applicability: T has guided strategic action or control, such as in science-based technology

Question. Again, how tractable are these virtues?

5 Aesthetic virtues

Beauty: T evokes aesthetic pleasure in properly functioning and sufficiently informed persons.

Question. *Why* should beauty have anything to do with truth?

Perhaps it can be connected to the following theoretical virtues:

Unification: T explains more kinds of facts than rivals with the same amount of theoretical content.

(**Note:** This is different from having a **broad scope!**)

Question. How to measure theoretical content?

Question. Is there a link between explanation and unification? (e.g. as in Kitcher's **argument patterns**)

Simplicity: T explains the same facts as rivals, but with less theoretical content.

Question. How to measure simplicity?

e.g. number of axioms, complexity of axioms, number of theoretical posits, number of *kinds* of theoretical posits.

Often this is illustrated with the example of **curve fitting**, simplicity can be in tension with evidential accuracy.

Question. Is there agreement on simplicity across the sciences?

(**Note:** Certainly not in mathematics—see work by Inglis and Aberdein.)

Question. Does it matter whether there is?

Question. If there is agreement, is it important if it's sociological?

Question. Is there a relationship between simplicity and unification?

6 Further questions

Question. Should we view (some) the anti-realists as advocating a kind of suspension of judgement?

Question. Do we really follow theoretical virtues in theory choice?

Question. Can a theory can be worse along all dimensions, but still represent an advance? (e.g. when observation hasn't caught up yet.)

Question. How context sensitive should the virtues be?

Question. Does this link to pluralism?