LSE Course Ph201 Scientific Method
BA Intercollegiate Philosophy Degree:
Philosophy of Science

Theory, Science and Realism

Lecture Notes

Dr Stathis Psillos © 1995

I. Theories of Scientific Growth

1. Popper’s Deductivism

Popper’s theory of science can be best described as anti-inductivist and falsificationist.

Anti-inductivism: Popper start with Hume’s Problem of Induction: a) induction is demonstrably a logically invalid mode of inference (induction is not deduction); b) any other way to justify induction, i.e., to show that we can rationally form our expectations about the future based on our past experiences, is bound to be circular. For such an attempted justification requires that the future will resemble the past. (It requires that since in the past, occurrences of **A** had always been followed by occurrences of **B**, so too in the future, occurrences of **A** will be followed by occurrences of **B**.) The only way to ‘justify’ this presupposition is to show that induction is successful. That is, that whenever we made a prediction that the next A was going to be B based on the past experience that so far all As had been Bs, this prediction was verified. But the evidence we’ve got about the success of induction is all to do with the past (past applications of inductive inferences) and hence it can apply to the future applications only if we show that the future will resemble the past. Clearly, we are now moving in circles! For Hume induction is irrational. Yet it is a psychologically indispensable ‘habit’ or ‘custom’.

But Popper went beyond Hume: (c) Induction is not just irrational. Induction is a myth. Real science doesn’t (and shouldn’t) employ induction.

Scientific hypotheses are normally universal generalisations of the form “All Ravens are Black”, or “All pendula are such that their period satisfies the equation $T=2\pi\ (l/g)^{1/2}$”, or “All atoms are composed of a nucleus and orbiting electrons” . Clearly, scientific hypotheses are never proved by evidence (i.e., the evidence never entails any (interesting) scientific hypothesis). But, as we’ve already seen, this is OK, if we can nonethesess show that evidence can confirm a hypothesis.
For Popper (d) scientific hypotheses are never confirmed by evidence. Observing positive instances of a hypothesis never raises its probability. To think otherwise is to think inductively, i.e., to think that scientists can learn from experience. Observing positive instances of a universal generalisation and then thinking that the more the positive instances, the more confirmation is accrued to the generalisation, i.e., the greater the confidence that the generalisation is true, is typical of inductive learning. For Popper inductive learning is a myth.

Falsificationism is the view that we needn't despair if inductivism fails! Scientific theories can be still falsified by the evidence. (There is a clear asymmetry between verification and falsification.) Testing scientific hypotheses is a deductive procedure. In particular, tests are attempts to refute a hypothesis.

Scientific hypotheses normally make claims about entities and processes that are not directly accessible in experience (unobserved ravens as well as unobservable electrons, light-waves etc.). How can these claims be tested? These hypotheses, together with initial conditions, logically entail several observational consequences (i.e., claims about things that can be directly observed (e.g., illuminated dots on a screen, interference fringes, pointer-readings, deflections of needles in measuring instruments etc.) [An observational consequence is a statement whose truth or falsity can be established by making observations.] These consequences (or predictions) are then checked by observations and experiments. If they turn out to be false (i.e., if the given prediction is not verified), then the hypothesis under test is falsified. But, if the predicted observational consequence obtains, the hypothesis from which it was drawn gets corroborated. (Be careful here: This sounds like the hypothetico-deductive theory of confirmation. But, on Popper's view, hypotheses are never confirmed. If the observation does not falsify the hypothesis, then the hypothesis does not become probable. It becomes corroborated. Corroborated is a hypothesis that i) has not yet been refuted and ii) has stood up severe tests (i.e., attempts at refutation).

So Popper's method of testing can be summarised as follows:

H (hypothesis to be tested)
C (Initial Conditions)

∴ **O (Observational consequence; prediction)**

If O, then H is unrefuted (corroborated); if not-O, then H gets falsified

Example: Boyle's law: Pressure * Volume=constant (at constant temperature) (ie. $P_i * V_i = P_f * V_f$)

H: Boyle's Law
C: Initial Volume=1m^3; initial pressure=1atm; final pressure=.5atm;

∴ O: Final volume 2m^3

(Measure the final volume to see whether the hypothesis is corroborated)

“It should be noticed that a positive decision can only temporarily support the theory, for subsequent negative decisions may always overthrow it. So long as a theory withstands detailed and severed tests and is not subsequently superseded by another theory in the course of scientific progress, we may say that ‘it has proved its mettle’ or that it is ‘corroborated’ by past experience” (Popper, LScD)

Incidentally, falsifiability is the criterion Popper uses to demarcate science from non-science. (**N.B.** Falsifiability is different from falsification.) Scientific theories are falsifiable in that they entail observational predictions (what Popper has called ‘potential falsifiers’) that can be tested and either corroborate or falsify the theory. On the contrary, non-scientific claims do not have potential falsifiers.

Aren’t there problems with Popper’s views? Severe problems. I shall just skip Popper’s critique of inductivism. I shall mention only one point. Hume criticised the rationality of induction but his criticism had a hidden premiss, viz., that in order to be able to form rational beliefs by means of induction we must first justify induction, that is, roughly, we must first establish that inductive reasoning generates true predictions. This is a substantive epistemological premiss (in fact a theory of justification) that has been recently contested by many philosophers.

If we understand corroboration à la Popper, then we face the following problem. Suppose we want to get a satelite into orbit. We are inclined to say that we’ll employ Newton’s theory. Why? Because it’s well-confirmed from past successful applications. This is an inductive move. Can Popper recommend the same action based on his notion of corroboration? That’s what he suggests, but notice that ‘corroboration’ is only to do with the past performances of the theory. It says nothing about how the theory is going to fare in a future application. If it made it likely that the theory would be successful in a future application, then it would clearly be akin to the inductive notion of confirmation. But Popper’s ‘corroboration’ is supposed to avoid the inductive leap from past successes to likely future successes. So Popper can recommend action on the basis of the best corroborated theory only if he makes an inductive leap. If you thought that at any stage there is only one corroborated theory available and that’s only rational to act on its basis, think again. Take Newton’s theory (NT). You can easily construct another theory NNT which says the following: ‘Everything in the universe is as Newton’s theory says until midday tomorrow but from then on the universal gravitation law will beome an inverse cube law.' Clearly, until midday tomorrow NT and NNT are equally corroborated. Which theory shall we use to get a rocket into orbit tomorrow afternoon? (Try to justify your answer.)

But is falsificationism immune to criticism? As Pierre Duhem first showed, no theory can generate any observational predictions without the use of several auxiliary assumptions. (In the previous example, the auxiliry assumption is that the temperature is constant.) But if the prediction is not fulfilled, then the only thing we can logically infer is that either the auxiliaries or the theory is false. (Logic alone cannot distribute blame to the premisses of an argument with a false conclusion.) Which means (just reflect on it for a minute) that as far as the logic of the situation is concerned, we can always attribute falsity to the auxiliaries and hold on to the theory come what may. (So, if the result is anything other than $2m^3$, one can and argue that either, after all, the temperature wasn't constant or that the measuring instruments are faulty.) So, the deductive pattern above is incomplete. The correct pattern is:

H (hypothesis to be tested)
C (Initial Conditions)
A (Auxiliary Assumptions)

$\therefore$ **O (Observational consequence; prediction)**

Which means that: If O, then the conjunction (H&C&A) gets corroborated. But can Popper distribute the praise? If, on the other hand, not-O, then the conjunction (H&C&A) gets falsified. But can Popper distribute the blame? (Notice that an inductivist can argue that, for instance, since Boyle's law has been confirmed in so many different circumstances and ways, then it's prudent to blame the failure to the auxiliaries. Conversely, if the auxiliaries come from background theories with a high degree of confirmation, the inductivist can blame the failure to the hypothesis. There is no cut-and-tried rule, but the inductivist can argue his case. Popper, on the contrary, cannot use the 'degree of corroboration' in a similar way for, as we saw, 'corroboration' applies only to past performances. But also this: Popper says that one shouldn't be allowed to use ad hoc auxiliary assumptions in order to save a hypothesis, where an auxiliary assumption is ad hoc if it's not independently testable. This is a promising move, but it faces two problems: a) how exactly to understand independent testability in a non-inductive way and b) sometimes, initially ad hoc hypotheses become testable. (You may discuss this in class).) If it's in principle possible to hold on to H come what may and blame the auxiliaries, then no hypothesis is, strictly speaking, falsifiable. And since we think that, in practice, hypotheses get in fact falsified by evidence, what we need are ways to distribute degrees of confirmation between theories and auxiliries so that it be possible to say that sometimes it's good to stick with the hypothesis whereas some other times it's good to abandon or modify it.

If scientific hypotheses are not 'induced' from experience (i.e., if they are not, typically, generalisations of observed facts) how are they arrived at? For Popper this is not a philosophically interesting question. It's a psychological rather than logical question. (As the expression goes, it belongs to the context of discovery rather than the context of justification.) Hypotheses may be arrived at by any means whatever (sleepwalking, guesswork, coin-tossing, serious thought etc.). Hypotheses are generally conjectured. The really philosophical issue is what happens after they have been suggested. How do they relate to the evidence? Are they testable? Are they refutable and if so, are they refuted by the evidence? This is, in a nutshell, Popper's methodology of conjectures and refutations:

1. Scientists stumble over some empirical problem.
2. A theory is proposed (conjectured) as an attempt to solve the problem ('tentative solution').
3. The theory is tested by attempted refutations ('error elimination').
4. If the theory is refuted, then a new theory is conjectured in order to solve the new problems. If the theory is corroborated, then it is tentatively accepted. (But it's not established, justified, probable and the like. It's just unrefuted.)

For example, Newton's theory of universal gravitation—Popper thinks—was conjectured in an attempt to solve the problem of deriving and explaining Kepler's laws and Galileo's laws. The theory solves this problem, but when it's further tested (attempted refutations) it failed to solve the problem of the haphazard motion of Mercury's Perihelion. A new theory is conjectured (Einstein's General Theory of Relativity) which attempts to solve this problem.

So, a scientific theory is abandoned when it clashes with observations and then a new theory is conjectured. (To be sure, the theory gets falsified only when we discover a reproducible effect that refutes the theory. This is what Popper called a 'falsifying hypothesis'.) This picture of theory-change can be best characterised as 'revolution in permanence'. The occurrence of scientific revolutions, in the sense of radical change of theory, is actually what constitutes progress in the Popperian theory of science. (Notice, in the Popperian scheme, there is no learning from experience. Experience does not show scientists how to modify their theories in order to cope with the anomalies; rather, experience only falsifies theories and, therefore suggests that another theory is in order.) But what makes this process of radical theory-change *rational*? And what guarantees the *objectivity* of theory-change in science?

Popper’s answer: a) rationality (in science and in general) is only a matter of attitude towards one’s own theories and views. It is the critical discussion of one’s own pet theory, its subjection to severe tests, its attempted refutation and, should it clash with observations, its elimination that renders theory-change rational. “As to the *rationality of science*, this is simply the rationality of critical discussion” (M of F).

But is this enough? Clearly, in order to judge progress and safeguard objectivity amidst a state of permanent succession of theories, one needs ways to compare the abandoned theory with the successor one and show that the latter is doing, in some sense, better than the former. Popper has to inject a certain dose of ‘conservativeness’ to the ‘Revolution in Permanence’ thesis. He suggests that b) the successor theory, however revolutionary, should always be able to explain fully and improve on the successes of its predecessor. In other words, the new theory should preserve the true empirical content of its predessesor and be able to use its own new conceptual resources to fully explain the succeses of its predecessor.

The general problem with Popper’s view is that it doesn’t really do justice to the way in which scientific theories are developed and, especially, modified. It is certainly true that scientific hypotheses are devised in order to explain some phenomena (Popper’s ‘problems’). But it’s not true that scientific theories are conjectured in vacuum, as Popper’s view implies. Rather, new scientific hypotheses are developed in the light of existing *background theories* and more general principes and symmetries in an attempt to bring the phenomena under the scope of the theory. Most of the time the existing background theories need to be considerably modified in order to be able to explain the new phenomena, and sometimes the necessary modifications are so radical in character that they lead to the abandonment of the old theoretical framework. This may be characterised as a revolution. But the occurrence of such a radical change is much more rare than Popper’s theory implies. Similarly, theories are not abandoned when the first contradictory observation is in sight. It’s much more accurate to say that a theory gets abandoned only when (a) there are persistent anomalies that cannot be successfully accommodated within its scope and (b) there is another (independently motivated) theory that takes care of these anomalies and accounts for the well-confirmed content of the old theory. (These issues will crop up again when we discuss Kuhn’s and Lakatos’s theories of science.)

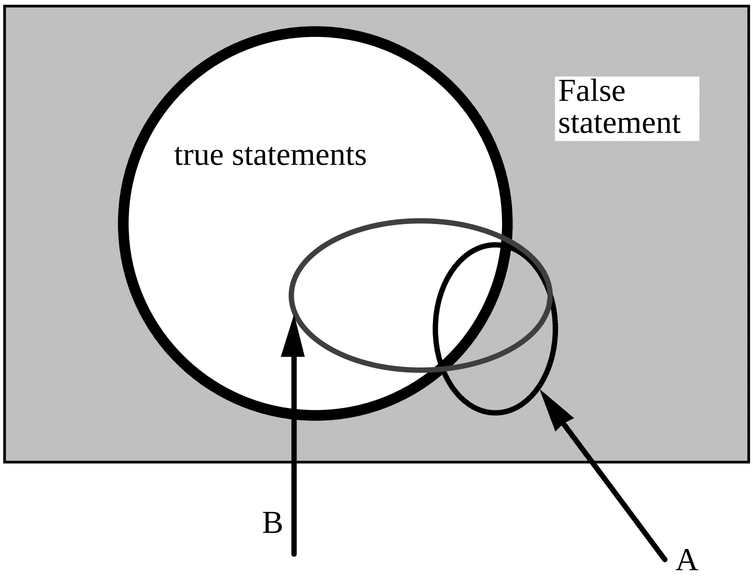
Where does this process of successive Popperian revolutions lead? Is the aim of science to produce unrefuted theories? And if all theories end up refuted by the evidence, is there any way to prefer one theory to another? Is there any way in which we can tell which theory is best to adopt? Notice that this is a serious problem for Popper’s theory of science, given that he thinks there is no way in which the evidence can (inductively) support a theory more than another. If we abandon induction we can never have any reason to believe in the truth of any contingent statement and similarly we can never have any reason to believe that one statement is more probable than another. Popper believes, however, that ultimately the aim of science is truth. (The concept of truth has been explicated by Tarski. For a brief but informative account of this theory, you may look at the appendix of Boyd’s piece ‘Confirmation, Semantics and the Interpretation of Scientific Theories’.) Does it make sense to strive for an aim such that we have, in principle, no way to say when it is realised? Even if truth is a concept well understood (a là Tarski), if Popper’s anti-inductivism is right then we have no criterion for truth (as, for instance strength of empirical support might have been for an inductivist). Worse than that, we have no way to say when it is, even in principle, reasonable to believe that a certain empirical statement is more likely to be true than another. How does Popper try to cope with this?

Popper suggests (quite correctly, I think) that the aim of science should be the development of theories with high verisimilitude (likeness to truth). Strictly speaking, all existing scientific theories are false. However, they may be closer to the truth than their predecessors and this surely constitutes progress. Besides, this can give us ways to compare theories and opt for the theory with higher verisimilitude. If, as science grows, we move on to theories with higher verisimilitude, then there is a clear sense in which this process takes us closer to the truth (although, at any given point, we may not know how close to the truth we are—discuss this issue in class, if you find it surprising). This is a great idea. But the difficulty lies with the explication of the notion of verisimilitude. Obviously, an explication will not be adequate if it turns out that any two false theories are equally far from the truth. But this is exactly what happens with Popper’s explication of verisimilitude. Let’s see how this happens.

Popper’s definition:

A is less verisimilar than B iff the truth-content of A is less than the truth-content of B and the falsity-content of B is less than or equal to the falsity-content of A; or the truth-content of A is less than or equal to the truth-content of B and the falsity-content of B is less than the falsity-content of A.

(The content of a theory T is the set of all statements derivable from the theory (logical consequences of T). The truth-content T_T is the set of all true consequences of T and the falsity-content T_F of T is the set of all false consequences of T.)



But Popper's definition is flawed. (This was proved independently by Tichy (1974) and Miller (1974).) **Proof**: Assume A and B are both false and distinct theories. Popper's definition: A $<_V$ B iff (i) $A_T \wp B_T$ and $B_F \prod A_F$ or (ii) $A_T \prod B_T$ and $B_F \wp A_F$.

Assume $A_T \wp B_T$ and $B_F \prod A_F$. Then, B has at least an extra true consequence, say q (i.e., $q \square B_T$). Clearly there are some falsehoods common to both A and B (given that B_F is contained in A_F and is non empty). Take any of these false consequence common to A and B, say p (i.e., $p \square B_F$ and $p \square A_F$). Then p&q is a false consequence of B (i.e., $p\&q \square B_F$) but clearly not of A (i.e., $p\&q \square A_F$). Hence, contrary to our assumption, it's not the case that $B_F \prod A_F$. (There is a false consequence of B which is not a false consequence of A.)

Assume $A_T \prod B_T$ and $B_F \wp A_F$. Then, A has at least an extra false consequence, say r (i.e., $r \square A_F$). Take any false consequence of A, say k (i.e., $k \square A_F$) which is also a false consequence of B. Then clearly, $k \varnothing r$ is a true consequence of A (i.e., $k \varnothing r \square A_T$) but not of B (i.e., $k \varnothing r \square B_T$). Hence, contrary to our assumption, it's not the case that $A_T \prod B_T$. (There is a true consequence of A which is not a true consequence of B.)

The thrust of this counterexample is this. In Popper's approach, it turns out that all false theories are equally distant from the truth. If we try to get a more verisimilar theory B from a false theory A by adding more truths to A, we also add more falsehoods. Similarly, if we try to get a more verisimilar theory B from a false theory A by subtracting falsehoods from A, we also subtract truths. The intuitive weakness of Popper's approach is that it ties verisimilitude to how much the theory says (i.e., how many true consequences it entails) and not to what the theory says.

2. Kuhn's Image of Science

Kuhn's theory of science can be seen as the outcome of two inputs: a) a reflection on the actual scientific practice as well as the actual historical development and succession of scientific theories and b) a reaction to what was perceived to be the dominant logical empiricist and Popperian images of scientific growth: a progressive and cumulative process that is governed by specific rules as to how evidence relates to theory and leads to theories with ever-increasing empirical support or verisimilitude.

Kuhn's own account of the dynamics of scientific growth can be schematically represented as follows:

paradigm ∅ normal science ∅ puzzle-solving ∅ anomaly ∅ crisis ∅ revolution ∅ new paradigm ∅ ... .

A coarse-grained explication of this schema is as follows: The emergence of a scientific discipline (the move from pre-science to science, if you like) is characterised by the adoption of a paradigm, which can be broadly characterised as the body of theories as well as values and methods they share. A scientific community is created around the given paradigm. The paradigm is faithfully adopted by the community. It guides the training of the new generations of scientists, and the research into new phenomena. A long period of 'normal science' emerges, in which scientists attempt to apply, develop, explore, tease out the consequences of the paradigm. The paradigm is not under test or scrutiny. Rather, it is used to offer guidance to the attempted solution of problems. This activity is akin to puzzle-solving, in that scientists follow the rules laid out by the paradigm in order to a) characterise which problems are solvable and b) solve them. (Unsolvable problems are set aside.) Normal science articulates the paradigm further, by realising its potential to solve problems and to get applied to different areas. This rule-bound (as we shall see later, it is best characterised as exemplar-bound) activity that characterises normal science goes on until an anomaly appears. An anomaly is a problem that falls under the scope of the paradigm, it is supposed to be solvable but persistently resists solution. The emergence of anomalies signifies a serious decline in the puzzle-solving efficacy of the paradigm. The community enters a stage of crisis which is ultimately resolved by a revolutionary transition (paradigm shift) from the old paradigm to a new one. The new paradigm employs a different conceptual framework, and sets new problems as well as rules for their solutions. A new period of normal science emerges.

The crucial element in this schema is the paradigm shift. According to Kuhn this is definitely not rule-governed. It's nothing to do with degrees of confirmation or conclusive refutations. Nor is it a slow transition from one paradigm to the other. Nor is it a process of (even a radical) modification of the old paradigm. Rather, it's an abrupt change in which the new paradigm completely replaces the old one. The new paradigm gets accepted not because of superior arguments in its favour, but rather because of its proponents' powers of rhetoric and persuasion. Paradigm shift is a revolution, pretty much like a social revolution, in which "a mode of community life" is chosen (SSR, 94). I shall say a bit more about revolutions below. Now let me explain (and criticise) Kuhn's basic concepts in some detail:

Paradigm: Kuhn has no uniform usage of this term. But the dominant characteristics of a paradigm are the following: a) "the entire constellation of beliefs, values, techniques, and so on shared by the members of a given community"; b) "the concrete puzzle-solutions which, employed as models or examples, can replace explicit rules as a basis for the solution of the remaining puzzles of normal science" (SSR, 175). So, the paradigm plays two broad roles:

a) it stands for the whole network of theories, beliefs, values, methods, objectives, professional and educational structure of a scientific community. Thus, it provides the bond of this community, characterises its world-view and guides its research.

b) at a more concrete level, it stands for a set of explicit guides to action (what Kuhn sometimes calls 'rules'). These are primarily understood as a set of exemplars (models, model-solutions, analogies etc.) that scientists are supposed to use in their attempt to solve problems and further articulate the paradigm.

In order to clarify this dual role of the paradigm, Kuhn later on replaced this concept by two others: disciplinary matrix and exemplars. The disciplinary matrix is now given a more precise account. It includes:

a) ‘symbolic generalisations’: that is the generalisations that a scientific community accepts as characterising the laws of nature or the fundamental equations of theories.
b) ‘models’: that is the set of heuristic devices and analogies that the theories make available for the description of phenomena. (But Kuhn also includes in models the most general convictions of the scientific community concerning the fundamental characteristics (constituents) of the world.)
c)‘values’: that is, all features that are used for the evaluation of scientific theories. Kuhn has identified five characteristics of a “good scientific theory” (1977, 321-322).
Accuracy: the theory should be in good agreement with the results of observations and experiments.
Consistency: the theory should be free from internal contradictions, but also in harmony with other accepted theories.
Broad Scope: the theory should explain disparate phenomena, especially phenomena that extend beyond those “it was initially designed to explain”.
Simplicity: the theory should “bring order to phenomena that in its absence would be individually isolated and, as a set, confused”. (Simplicity is understood as minimising the number of independently accepted hypotheses.)
Fruitfulness: the theory should be able to guide new research, in particular to “disclose new phenomena or previously unnoted relationships among those already known”.

For Kuhn, however, these are not necessarily trans-paradigm values. Nor does he think that they can be used to rationally adjudicate between competing paradigms. (For a detailed account of the role of these values in theory choice, one should look at Kuhn’s (1977).)

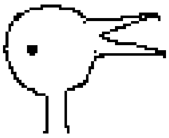
Exemplars, on the other hand, are conceived of as exemplary problem solutions (literally, of the kind you find in any physics textbook). Exemplars specify, or concretise, the meaning of the fundamental concepts of the paradigm (e.g., the meaning of ‘force’ is further specified by means of exemplars that employ certain force-functions, e.g., Hooke’s law.) Exemplars are also used for the identification of new research problems and are normally employed for the solution of such problems (e.g., we use the exemplar of harmonic oscillator to study the behaviour of new periodic phenomena).

There is a certain sense in which a Kuhnian paradigm (in its broader sense) defines a world. And there is a certain sense, as Kuhn says, in which when a new paradigm is adopted, the world changes. “The proponents of competing paradigms practice their trades in different worlds.... Practicing in different worlds, the two groups of scientists see different things when they look from the same point of view in the same direction” (SSR, 150). And similarly: “In learning to see oxygen, however, Lavoisier also had to change his view of many other more familiar substances. He had, for example, to see a compound ore where Priestley and his contemporaries had seen an elementary earth, and there were other such changes besides. At the very least, as a result of discovering oxygen, Lavoisier saw nature differently. And in the absence of some recourse to that hypothetical fixed nature that he ‘saw differently’, the principle of economy will urge us to say that after discovering oxygen Lavoisier worked in a different world” (SSR, 118). We must take these expressions literally and not metaphorically.

The best way to understand Kuhn’s general philosophical perspective is to say that Kuhn’s underlying philosophy is relativised neo-kantianism. It’s neo-kantianism because he thinks there is a distinction between the world-in-itself, which is epistemically inaccessible to cognizers, and the phenomenal world, which is in fact constituted by the cognizers’ concepts and categories (paradigms, if you like), and is therefore epistemically accessible. But Kuhn’s neo-kantianism is relativised because Kuhn thinks that there is a plurality of phenomenal worlds, each being dependent on, or constituted by some community’s paradigm. So, different paradigms create different worlds. This needs some reflection in order to be properly understood. But you may think of it as follows: (sketchy) The paradigm imposes a structure on the world of appearances (in the world of stimuli with which our sensory receptors are bombarded). It carves up this world in ‘natural kinds’. This is how a phenomenal world is ‘created’. But different paradigms carve up the world of appearances in different networks of natural kinds. So, for instance, Priestley’s paradigm carves up the world in such a way that it contains dephlogisticated air whereas Lavoisier’s paradigm carves it up is such a way that it includes oxygen. It is in this sense that Priestley and Lavoisier inhabit different worlds--but notice this is not just a metaphor. Try to explore this issue in class. For more on this look at Hoyningen-Huene’s book, chapter 2.)

These thoughts lead us naturally to the idea of incommensurability between competing paradigms. (Kuhn says that the ‘most fundamental aspect of ... incommensurability” is the fact that “the proponents of competing paradigms practice their trades in different worlds” (SSR, 150)). Kuhnian paradigms are not inconsistent. (Claiming that **p** is inconsistent with **q** requires i) that **p** can be compared with **q** in terms of meaning and ii) showing that if **p** is true, then **q** is false, and conversely.) Competing paradigms cannot even be compared. They are incommensurable: there is no way in which we can systematically translate talk in terms of one paradigm into talk in terms of the other paradigm; competing paradigms are not intertranslatable. According to Kuhn, the paradigm as a whole defines the meaning of the terms employed in it. When there is a paradigm shift, there is no reason to suppose that the meaning of terms remains the same. Nor can we suppose that if a term occurs in two different paradigms it refers to the same 'things' in the world. In paradigm shift both the intention (meaning) and the extension (reference) of terms changes. So, for instance in the passing from the Ptolemaic to the Copernican paradigm the meaning and reference of the term 'planet' changed. And similarly for the term 'mass' in the transition from the Newtonian paradigm to the Einsteinian one. (We'll discuss the concept of meaning incommensurability in more detail in the next lecture.)

As I said already a Kuhnian revolution is an abrupt procedure in which a new paradigm is adopted. Kuhn has used an example from gestalt-psychology to illustrate the situation. This is the famous duck-rabbit case.



But whereas in ordinary gestalt cases, the subject can go back and forth between the image of the duck and the image of the rabbit, the scientists cannot do that. The shift to the new paradigm 'transports them to a different world'. It is precisely the incommensurability of competing paradigms that accounts for the alleged impossibility to understand both paradigms. These claims have led many philosophers to characterise Kuhn's theory of science as relativist and irrationalist. (In his later writings Kuhn has attempted to disown these characterisations.)

Normal Science: We said already that normal science is when the paradigm is further developed and articulated by means of puzzle-solving activity. Kuhn talks of normal science as "an attempt to force nature into the preformed and relatively inflexible box that the paradigm supplies" (SSR, 24). More specifically, during normal science no innovation is intended or expected. Puzzle-solving consists in refining the paradigm not in challenging it. Similarly, during normal science, the paradigm is not tested. INo attempt is made either to confirm the paradigm or to refute it. Within normal science, neither the conceptual framework employed in exemplary solutions nor the generalisations, models, values, commitments etc are under review or dispute. As you can see, Kuhn's normal science is fairly dogmatic.

Is there any sense in which there is progress in normal science? There is a trivial sense in which normal science can progress, viz., when it successfully solves more problems and puzzles from the point of view of the paradigm. Then, one can say that normal science increases our knowledge of the world. But which world? The paradigm's own phenomenal world, that is the world 'created', or shaped by the paradigm. This is, I think, only self-indulging progress. Think of what happens when the paradigm is abandoned. If the new paradigm defines a new phenomenal world, the old paradigm has clearly made no progress with respect to the new world. Isn't then all talk of progress in vain? (Kuhn himself more or less confirmed that when he said "With respect to normal science, then, part of the answer to the problem of progress lies simply in the eye of the beholder" (SSR, 162-163). If questions of progress can only be asked in relation to a certain paradigm, then is there any sense in which we can compare two paradigms and ask whether one is more progressive than the other?

What happens to claims of progress if paradigms are incommensurable? Isn't there any way in which one paradigm can be said to be more progressive than another? If the old set of problems (exemplars, etc.) is replaced by a new one defined within the new paradigm how can paradigms be compared with respect to their progressiveness? Clearly, judgements of progress require a reference point, but Kuhn doesn't offer us any. If some problems remained invariant under paradigm shift, then their solution would clearly constitute a measure of progress. In fact Kuhn suggests that a) the new paradigm must be able to solve "some outstanding and generally recognised problem that can be met in no other way" and b) the new paradigm must "promise to preserve a relatively large part of the problem-solving ability that has accrued to science through its predecessors" (SSR, 168). But it should be clear that this would require that the old and new paradigms be commensurable. For otherwise, there can be no way to identify a problem as being the same in both paradigms, nor to preserve the existing problem-solving ability. In fact Kuhn has more recently reviewed his position concerning incommensurability. To be sure, it is still there but it's local rather than global. It occurs only when the competing theories have locally different taxonomies of natural kinds (what, more or less, Kuhn calls different lexical structures).

Is theory-change in science irrational? Notice that if theories (or paradigms) are incomparable, then surely any choice between them would have to be fundamentally irrational. There could be no way to compare them and hence to offer grounded reasons to prefer the one over the other. Kuhn makes the best use of Duhem's thesis that no theory can, strictly, speaking be falsified: a theory is never falsified; it is only replaced by another theory when and if this becomes available. But still there should be ways to guide the choice between the old and the new theory. (This is in fact what Lakatos tried to show as we shall see in the next section.) But even without entering into the details of the problem we can say this. Take Kuhn’s values of the disciplinary matrix: accuracy, consistency, scope, simplicity, fruitfulness. Why can’t we use these values to ground our judgements concerning theory-choice? Kuhn is right to note that there is no algorithmic procedure to decide which theory fares better vis-à-vis the satisfaction of these values. (Kuhn says that the criteria of theory choice “function not as rules, which determine choice, but as values which influence it”.) Even within the same paradigm, different individual evaluators that employ the same value system may come to different conclusions. But why should rational judgement be conflated with algorithmic judgement? All that follows from Kuhn’s claims is that values can only incompletely determine theory choice. However, it does not follow that they cannot guide scientists to form a rational preference for one theory over the other.

3. Lakatos’s Progressivism

Lakatos’s main project may be characterised as follows: to combine Popper’s and Kuhn’s images of science in one model of theory-change that preserves progress and rationality while avoiding Popper’s naive falsificationism and doing justice to the actual history of radical theory-change in science. To this end, Lakatos developed the Methodology of Scientific Research Programme (MSRP).

The Research Programme (RP) is the unit by which the nature and direction of scientific growth is analysed and appraised. Lakatos rejects the view that progress should be examined vis-à-vis competing individual theories. Rather, progress can only be examined in relation to competing sequences of theories that constitute different research programmes.

Which sequence of theories constitutes a particular research programme? A research programme is characterised by three constituents: a) the hard-core, b) the negative heuristic and c) the positive heuristic.

The hard-core consists of all those theoretical hypotheses that any theory which belongs to this RP must share. These are the most central hypotheses of the RP, the ones that the advocates of the RP decide to hold irrefutable. So, when the RP faces anomalies, the hypotheses of the hard-core are those that are deemed immune to revision. If, in the face of anomalies, the RP requires revisions, the corrective moves will all be directed away from the hard-core and towards other hypotheses of the RP (e.g., several auxiliary assumptions, low-level theoretical hypotheses, initial conditions etc.). This methodological decision to protect the hard-core from revision constitutes the negative heuristic of the RP. For instance, the hard-core of Newtonian Mechanics (NM) is the three laws of motion together with the law of gravitation. When the orbit of the newly discovered planet Uranus (Herschel 1871) was shown not to conform with the predictions of the Newtonian hard-core, the negative heuristic dictated that the relevant modifications should be directed towards some auxiliary assumption, viz., the assumption that H: 'The trajectory of Uranus is unperturbed by the gravitational attraction of other nearby planets'. H was modified so that the trajectory of Uranus was re-calculated, now taking into account the gravitational attraction by Jupiter and Saturn (Pierre Laplace; Alexis Bouvar c.1820). When the new prediction was again not fulfilled, the negative heuristic dictated that the further modifications necessary should still be directed towards the auxiliaries. The new modification was the hypothesis H^*:'There is a hitherto unobserved planet—Neptune—whose motion perturbs that of Uranus'. That's (more or less) how Neptune was discovered (Adams; Leverrier c. 1845)! (Can Popper's theory satisfactorily account for this historical episode? Justify your answer.) This process of articulating suggestions as to how the RP will be developed, either in the face of anomalies or as an attempt to cover new phenomena, constitutes the positive heuristic of RP. In the case of NM the positive heuristic includes plans for increasingly accurate models of planetary motions. As Lakatos put it: the positive heuristic "defines problems, outlines the construction of a belt of auxiliary hypotheses, foresees anomalies and turns them victoriously into examples, all according to a preconceived plan". So, the positive heuristic creates a "protective belt" around the hard-core which absorbs all potential blows from anomalies.

Progress in science occurs when a progressive research programme supersedes another degenerating one. When this happen, the degenerating rival gets eliminated (or, "shelved"). A Research Programme is progressive is as long as "its theoretical growth anticipates its empirical growth", that is as long as it continues to predict novel facts some of which are corroborated. This constitutes a progressive problemshift. A Research Programme is stagnating if "its theoretical growth lags behind its empirical growth", that is if it does not predict novel facts but it only offers post hoc explanations of facts either discovered by chance or predicted by a rival research programme. This constitutes a degenerating problemshift. It's critical to note the emphasis that Lakatos put on the role of novel predictions in judging progressivess. Initially, the idea of a novel prediction was that the predicted phenomenon is hitherto unknown (a novel fact). This is a clear sign of progressiveness, because if the theory predicts a novel fact, then clearly the theory goes ahead of the experiment, sets tasks to the experiment, takes extra risks. In short, theoretical knowledge grows faster than empirical knowledge. But it is clear that theories are also supported by successfully explaining already known phenomena (cf. Einstein's General Theory of Relativity gained considerable support by explaining the anomalous perihelion of Mercury). How can MSRP account for this? Lakatos suggests that a RP can also gain support by explaining an already known phenomenon, provided that this phenomenon was not used in the construction of the theory. If this happened, the theory would be ad hoc with respect to this phenomenon. In fact, it wouldn't possibly fail to explain it.

Lakatos has recognised that the procedure of evaluation of competing research programmes is not always straightforward. In fact, he has conceded that it is difficult to decide when an RP is irrevocably degenerating and that it's possible for a degenerating programme to recover and stage a comeback. This creates a rather important problem for Lakatos's account. For how can we write off a research programme as degenerating given that if we wait long enough it may recover? And is it irrational to advocate a degenerating research programme? (Lakatos himself said: "One may rationally stick to a degenerating research programme until it is overtaken by a rival and even after" (1981, 122).) The problem is intensified if we take into account the following: as Hacking has noted, Lakatos's methodology is retroactive. It provides no way to tell which of two currently competing RPs is progressive. For even if one of them seems stagnated, it may stage an impressive comeback in the future. Lakatos's methodology can only be applied to past research programmes where with sufficient hindsight we can tell that they either have entered into a stage of degeneration (no novel predictions any more), or that they have exhibited progressive problemshift for a long period of time. But when is the hindsight sufficient enough to suggest that a certain RP should be "shelved" for good?

No matter what one thinks about the last point, there is no doubt that Lakatos's view of science overcomes the major difficulties faced by the theories of Popper and Kuhn. It's not naively falsificationist: a falsifying hypothesis falsifies a theory only after it has been satisfactorily explained by another theory. It makes progress dependent on the development of different theories at the same time. It acknowledges the impact of Duhem's thesis and accepts that falsification requires several methodological decisions. Similarly, Lakatos's view safeguards progress in science and avoids the perplexities of incommensurability. It also avoids the dogmatism of 'normal science' while accepting that until the RP is abandoned research is guided by the "hard-core"—something not totally alien to Kuhn's paradigm.

Study Questions

1. State carefully Popper’s definition of verisimilitude and show that, on this account, all false theories are equally distant from the truth.

2. Reflecting on the Miller-Tichy result, Popper suggested the following amendment to his definition of verisimilitude: B is more truthlike than A just in case the truth-content of A is less than the truth-content of B (i.e., $A_T \wp B_T$). Does this suggestion work? Justify your answer. (Hint: B clearly entails all truths that A entails. What about falsehoods? Use the Miller-Tichy result.)

3. Carefully explain how Duhem’s critique of falsificationism affects Popper’s position. Can Popper’s position be saved?

4. Analyse and critically examine Kuhn’s concepts of ‘paradigm’ and ‘normal science’.

5. “Though the world does not change with a change of paradigm, the scientist afterward works in a different world” (Kuhn, SSR, 121). Discuss.

6. Why does Kuhn say that the successes of the paradigm do not confirm it?

7. Explain the role of novel predictions in Lakatos’s account of progress in science.

References

Hacking, I. (1981) ‘Lakatos’s Philosophy of Science’ in I. Hacking (ed.) Scientific Revolutions, Oxford UP.

Kuhn, T.S. (1977) ‘Objectivity, Value Judgement and Theory Choice’, in The Essential Tension, Chicago UP, pp.320-351.

Lakatos, Imre (1981) ‘History of Science and its Rational Reconstructions’ in I. Hacking (ed.) Scientific Revolutions, Oxford UP.

Miller, D.W. (1974) ‘Popper’s Qualitative Theory of Verisimilitude’, British Journal for the Philosophy of Science, **25**, pp.166-177.

Oddie, G. (1986) Likeness to Truth, Dordrecht: D Reidel Publishing Company.

Tichy, P. (1974) ‘On Popper’s Definitions of Verisimilitude’, British Journal for the Philosophy of Science, **25**, pp.155-160.

II. Theory, Observation, and Theoretical Terms

5. Concept Empiricism and the Meaning of Theoretical Terms

Empiricists have always tied meaningful discourse with the possibility of some sort or another of experiential verification. Assertions are said to be meaningful iff they can be verified. This has been widely known as the verifiability theory of meaning. Moritz Schlick and Rudolf Carnap once thought that the meaning of a statement is given in the method of its verification. And Russell before them defended the Principle of Acquaintance: “Every proposition that we can understand must be composed of ingredients with which we are acquainted”. This family of views may be called ‘concept empiricism’: concepts must originate in experience. It then follows that the meaning of a word must be either directly given in experience (e.g., by means of ostention) or be specified by virtue of the meanings of words whose own meaning is directly given in experience. If a word fails to satisfy either of these conditions, then it turns out not to be meaningful (meaningless). What is it, anyway, to specify the meaning of a word **a** (the definiendum) in terms of the meanings of other words, say **b** and **c** (the definiens)? (Think of the word ‘bachelor’ and the words ‘unmarried’ and ‘man’.) The natural thought here is that this specification involves an explicit verbal definition of **a** in terms of **b** and **c** (roughly, ‘bachelor’ iff ‘unmarried man’). If the definiens are meaningful, so is the definiendum. Given an explicit definition of a word **a**, **a** can be systematically replaced in any (extensional) context by its definiens.

But scientific theories posit a host of unobservable entities and mechanisms that are not directly accessible in experience (electrons, fields, genes, atoms, valence, I.Q., social classes etc.). Scientifc discourse seems to be, at least partly, about such entities. If concept empiricism is true, is this discourse meaningless? If statements about these entities cannot be directly verified, are these statement meaningless? Can one possibly be any kind of realist regarding unobservables?

There is a straightforward, if implausible, answer to these questions. Semantic instrumentalists think that precisely because statements about unobservables are unverifiable, all theoretical talk is, strictly speaking, meaningless. Semantic instrumentalism claims that assertions involving theoretical terms are meaningful—or cognitively significant—insofar as they can be translated into assertions involving only observational terms. But insofar as they are not so translated, they are meaningless. This is an instrumentalist position for it suggests that scientific theories do not issue in commitments to unobservable entities. On this account, if some theoretical terms are definable by virtue of observable terms and predicates, then they are clearly dispensable (eliminable). At best, they are shorthands for complicated connections between observables. Theoretical discourse ends up being nothing but disguised talk about observables, and therefore it's ontologically innocuous. For it is no longer seen as referring to unobservable entities and processes, and, therefore it issues no commitments to them. But if theoretical terms are not so defined, then, on the semantic instrumentalist account, they are meaningless, their only value being instrumental: they help in an economic classification of observable phenomena.

Discussing in some detail Carnap's attempts to defend concept empiricism may help us see why the semantic instrumentalist idea has failed. For Carnap was probably the only philosopher to take seriously the challenge of examining whether theoretical terms can be explicitly defined by virtue of observational terms and predicates (which were taken to be independently meaningful). The failure of this project, as we shall see in some detail later on, made Carnap abandon the idea that the meaning of theoretical terms can and should be completely defined in virtue of observational ones. Carnap remained an empiricist but, roughly after 1936 and until the end of his life in 1970, his main preocuppation was to defend a weakened version of concept empiricism. As he put it: "As empiricists, we require the language of science to be restricted in a certain way; we require that descriptive predicates and hence synthetic statements are not to be admitted unless they have some connection with possible observations, a connection which has to be characterized in a suitable way" (1937, p.33). Theoretical terms were to be admitted as meaningful insofar as "they have some connection with possible observations", but, as we shall see the requirement of definability was abandoned.

6. Carnap’s Empiricist Adventures

Let me begin by discussing the concept of an observational term or predicate. Carnap suggested that a predicate P of a given language stands for an observable property or relation, (that P is an observational predicate), if a person can, under suitable circumstances, decide with the aid of observations alone whether or not an object belongs to the extension of this predicate. In other words, a (monadic) predicate P is observational if for a given object **b,** observations can decide between the atomic sentences ‘Pb’ and ‘¬Pb’ (1936, pp445-446). Analogously, a theoretical predicate is one that does not satisfy the foregoing condition.

Concerning observational predicates, it is important to note two things. First, the decision procedure is dissociated from the old positivist requirement of complete verification of either ‘Pb’ or ‘¬Pb’, in that it is no longer required that the truth-value of either of those atomic statements be infallibly established. That strict verification is not possible in the case of universally quantified statements, which normally express laws of nature, was quickly observed by Carnap and other empiricists, especially after Popper’s Logik der Forschung, in 1935. (This was to become The Logic of Scientific Discovery.) But Carnap and others also accepted that the difference between universally quantified statements and atomic, or singular, ones is only one of degree, even when the atomic sentences refer to observable state-of-affairs. Carnap argued that even if one amasses a considerable quantity of evidence (test-observations) which suggest the truth of an atomic sentence such as ‘a desktop computer is on the table’, it is still logically and theoretically possible that one is wrong about it, and that further test-observations could show this (cf. 1936, p.425).[1]

Second, the distinction between observable predicates and theoretical ones is not sharp and immuttable. Carnap says: “There is no sharp line between the observable and non-observable predicates because a person will be more or less able to decide a sentence quickly, ie he will be inclined after a certain period to accept the sentence. For the sake of simplicity we will here draw a sharp distinction between observable and non-observable predicates. But thus drawing an arbitrary line between observable and non-observable predicates in a field of continuous degrees of observability we partly determine in advance the possible answers to questions such as whether or not a certain predicate is observable by a given person” (1936, p.455). So, surprisingly enough, Carnap already in 1936

[1] This view is, roughly, the outcome of the so-called protocol-sentences debate: Carnap was, in fact persuaded by Otto Neurath that all observational statements are hypotheses wrt elementary experiences. Neurath suggested that no statement is immediately verifiable nor immune to revision.) The ideal of conclusive verification of an atomic sentence was then replaced by the weaker requirement of confirmation of either ‘Pb’ or ‘¬Pb’.

countenanced a view which many of critics in the 1960's used in an attempt to knock down the whole idea of the existence of a theory-independent observational language. The distinction that Carnap drew was one of convenience, for as he, I think correctly, stated "if confirmation is to be feasible at all, this process of referring back to other predicates must terminate at some point" (1936, p.456).

Let us now take these consideration on board and try to explore Carnap's attempts to specify the meaning of theoretical terms—"in a suitable way"—in virtue of observable terms.

6.1 Explicit definitions

Carnap first tried to see how far one can go in an attempt to introduce theoretical terms by means of explicit definitions. Those have the form:

$$\forall x\ (Qx \times (Sx \varnothing Rx)). \tag{D}$$

(D) says that the theoretical term Q applies to x if and only if, when x satisfies the test-condition S, then x shows the observable response R. So, for instance, an explicit definition of the theoretical term 'temperature' would be like this. An object **a** has temperature of c degrees centigrade if and only if the following condition is satisfied: if **a** is put in contact with a thermometer, then the thermometer shows c degrees on its scale. The conditional $(Sx\varnothing Rx)$ is what Carnap called 'scientific indicator'. Scientific indicators express observable states-of-affairs that are used as the definiens in the introduction of a term. Carnap hypothesised that "in principle there are indicators for all scientific states of affairs (1928, §49).[2]

If the conditional $(Sx\varnothing Rx)$ is understood as material implication, then you can easily see that $(Sx\varnothing Rx)$ is true even if the test-conditions S do not occur. So, explicit definitions turn out to be empty. Their intended connection with antecedently ascertainable test conditions is undercut. For instance, even if we do not put an object **a** in contact with a thermometer, it follows from the explicit definition of 'temperature' that the temperature of **a** is going to be c degrees centigrade (whatever this numerical value may be). In order to avoid this problem, the conditional $Sx\varnothing Rx$ must not be understood as material implication, but rather as strict implication, ie as asserting that the conditional is true only if the antecedent is true.

[2] The initial thought was that the definiens will ultimately involve only terms and predicates with reference to 'elementary experiences' (let's say 'sense-data'). But, as we hinted to already, Carnap soon abandoned this aim and took the class of material (middle-sized) objects as his reductive basis and their observable properties and relations (size, colour, shape, heavier than etc.) as his basic reductive concepts.

But, even so, there is another serious porblem: in scientific practice, an object is not supposed to have a property only when test-conditions S and the characteristic response R actually occur. For instance, bodies are taken to have masses, charges, temperatures and the like, even when these magnitudes are not being measured. But, the logical form of explicit definition makes it inevitable that attribution of physical magnitudes is meaningful when and only when the test conditions S and the characteristic response R obtain. In order to avoid this unhappy conclusion, explicit definitions must be understood as counter-factual assertions, ie as asserting that the object **a** has the property Q iff if **a** were to be subjected to test-conditions S, then **a** would manifest the characteristic response R. In other words, theoretical terms have to be understood on a par with dispositional terms. However, the introduction of dispositional terms, such as 'is soluble', 'is fragile' and the like, was a problem of its own. It required a prior understanding of the 'logic' of counter-factual conditionals. Now that all theoretical terms have to be ultimately understood on a par with dispositional terms, the problem at hand gets worse. An appeal to 'nomological statements which subsume counter-factuals such as 'if x were submerged in water, x would dissolve' under general laws such as 'For all x, if x is put in water, then x dissolves', would prima facie provide truth-conditions to counter-factual conditionals and a basis for introducing dispositional terms in general. But nomological statements express laws of nature, which provide truth-conditions for these statements. How exactly we should understand laws of nature is a problem in itself. But no matter what one takes the laws of nature to be, they are neither observable nor explicitly definable in terms of observables. At any rate, even if all these problems concerning explicit definitions were tractable, it is not certain at all that all theoretical terms which scientists considered perfectly meaningful can be explicitly defined. Terms such as 'magnetic field vector', 'world-line', 'gravitational potential', 'intelligence' etc., can not be effectively defined in terms of (D), even if (D) were unproblematic.

Carnap suggested that this project is futile as early as in 1936 (in his magnificent 'Testability and Meaning'). In a certain sense, this failutre marks the end of semantic instrumentalism proper. For if theoretical terms are not explicitly definable, then they cannot be dispensed with by semantic means. Well, one may say, so much the worse for them, since they thereby end up meaningless. Not quite so, however. If theoretical terms end up meaningless because assertions involving them cannot be strictly speaking verified, then so much the worse for the verification theory of meaning. As we saw Carnap, Popper and others all agreed that, strictly speaking, even assertions involving only observational terms and predicates cannot be verified. They can only be confirmed—athough Popper disagreed with the last point. But clearly we are not willing to say that all these assertions are meaningless.

6.2 Reduction sentences

Having abandoned the idea of explicit definability, Carnap did not thereby abandon the empiricist demand that theoretical terms be introduced by reference to observables. What changed was his attitude about the possibility of eliminating theoretical discourse, not his view that it is necessary to specify, as far as possible, the meaning of theoretical terms by reference to overt and intersubjectively specifiable observable conditions. As he put it: “Reducibility can be asserted, but not unrestricted possibility of elimination and re-translation” (1936, p464). Carnap now suggested the introduction of a theoretical term or predicate Q by means of the following reductive pair:

$$\forall x\ (S_1x \varnothing (R_1x \varnothing Qx))$$
$$\forall x\ (S_2x \varnothing (R_2x \varnothing \neg Qx)) \qquad \text{(RP)}$$

in which S_1, S_2 describe experimental (test-)conditions and R_1, R_2 describe characteristic responses, (possible experimental results). In the case that S_1+S_2 (+S) and R_1+¬R_2 (+R), the reduction pair (RP) assumes the form of the bilateral reductive sentence

$$(x)\ (Sx \varnothing (Qx \times Rx)). \qquad \text{(RS)}$$

So suppose that we want to introduce the term Q: ‘temperature’ by means of a reductive sentence. This will be: if the test-conditions S obtain (ie, if we put **a** in contact with a thermometer), then **a** has temperature of c degrees centigrade iff the characteristic response R obtains (ie, if the thermometer shows c degrees centigrades). The reductive introduction of theoretical terms by means of (RS) does not face the problems of explicit definitions. If an object **a** is not under the test-condition S, then the sentence (Sa ∅ (Qa × Ra)) becomes true (false antecedent, true conditional). But this implies nothing as to whether the object under consideration has or has not the property Q. However, the reduction sentence (RS) does not define the predicate Q. For although (RS) provides a necessary and a sufficient condition for Q, these two conditions do not coincide. (It is easy to see that (Sx ∅ (Qx × Rx) is analysed as follows: (Sx □ Rx) ∅ Qx and (Sx □ □Rx) ∅ □Qx. This means that all things which are S□R are also Q, ie the concept Q definitely applies to them, and all things that are Q are also □(S□□R). But since □(S□□R) is different from S□R, the foregoing procedure does not specify all and only things to which the concept Q applies.) Thus, the meaning of Q is not completely specified by virtue of observable predicates. At best, the reduction sentence gives “a conditional definition” of Q (Carnap, 1936, p443). It can only give a partial empirical significance to a theoretical term (but only for the cases in which the test-conditions are fulfilled). Carnap thought that, ideally, a term could finally be associated with a whole set of reduction sentences which specify, partly, empirical situations in which the term applies (cf. 1936, §9). Feigl called Carnap’s programme a generalised case of ‘if-thenism’. For, it amounts to the introduction of theoretical terms, e.g. ‘magnetic field’, by means of a set of ‘if...then’ statements which specify testable observational conditions for their applicability. But even when a set of reduction sentences is specified, the term is not eliminable. For no amount of reduction sentences, being mere conditional definitions, could explicitly define—and hence render dispensable—a theoretical term. (Remember, reduction sentences are not explicit defintions.) As Carl Hempel put it, theoretical terms exhibit an “openness of content”, which is such that one introductive chain, “however rich, will still leave room for additional partial interpretations of the term at hand” (1963, p.689).

Already in 1939, Carnap suggested that theoretical terms are in fact indispensable. His argument was the following:

(1) Without using theoretical terms it “is not possible to arrive (...) at a powerful and efficacious system of laws” (1939, p64). That is, without appealing to theoretical entities, it is impossible to formulate laws applying to a wide range of phenomena. (Carnap was careful to add that “this is an empirical fact, not a logical necessity”.) Laws which involve only observational terms cannot be comprehensive and exact. They always encounter exceptions and have to be modified or narrowed down. On the contrary, laws that involve theoretical terms manage to encompass and unify a wide range of phenomena. In one phrase, the first premiss says: no theoretical terms, no comprehensive laws. (Statistical mechanics, for instance, explains why real gases only approximately obey the equation of state PV=RT by taking into account the molecular structure of gases. Real gases obey the more complicated Van Der Waals’ law $(P+n^2a/V^2)\infty(V-nb)=nRT$, where b depends on the size of the molecules and a on the intermolecular forces.)
(2) Scientists manage to formulate comprehensive laws with the help of theoretical entities. Therefore, theoretical terms are indispensable.

6.3 The two-tier model

The recognition that reduction sentences can only offer a partial empirical meaning to theoretical terms and that theoretical terms are indispensable for the development of scientifc theories marked a major turn in the empiricist programme. For as Hempel once put it: “(O)nce we grant the conception of a partial experiential interpretation of scientific terms through a combination of stipulation and empirical law, it appears natural to remove the limitations imposed by Carnap upon the form of reduction sentences and introductive chains” (1963, p691). Indeed, if we cannot anyway reduce talk about theoretical terms to talk about observable behaviour, then what does it matter what logical form we use in order to specify the partial empirical meaning of theoretical terms? For instance, the discovery of a new regularity of the form (x) (Qx ∅ Px), where Q is a theoretical term and P is an observable predicate, can clearly be used to further specify the range of applications of Q, even though it does not have the form of a reduction sentence.

Elaborating on this thought, Carnap (1956) suggested a two-tier, or two-language, model of scientific theories. The language of science is split into two sub-languages: an observational language L_O which is completely interpreted and whose vocabulary V_O designates observables and a theoretical language L_T whose descriptive vocabulary V_T consists of theoretical terms. A scientific theory T is then characterised by means of a set of axioms Γ formulated in V_T (in fact, a theory is the set of all deductive consequences of the axioms) and a set of **correspondence rules** C which are mixed sentences connecting the theoretical vocabulary V_T with the obsrvational vocabulary V_O. The V_T-terms are, in general, implicitly defined (interpreted) by the set of postulates Γ. What the correspondence rules are supposed to do is: a) to provide a **partial empirical interpretation** (meaning) to the theoretical terms in V_T; and b) to specify the conditions of their applicability. A typical example of a correspondence rule is this: the theoretical term ‘mass’ is connected with the observable predicate ‘is heavier than’ by means of the rule C: ‘the mass of body **u** is greater than the mass of body **v** if **u** is heavier than **v**’. But Carnap now thinks that there is no need to specify a single correspondence rule for each theoretical term in V_T. All that is required is some correspondence rules for some terms. Then their role gets fused. Since all theoretical terms are connected with one another via the postulates Γ of T, they all get connected with the specified correspondence rules and they all get some partial empirical interpretation. (It’s worth noting that this image of scientific theories—sometimes called the ‘received view’—has been recently contested by many philosophers of science who advocate the so-called semantic view of theores.)

What is worth pointing out here is that Carnap did not advance the two-tier model as a way to define theoretical terms in virtue of observational ones. Clearly, the model cannot furnish such definitions. Carnap had already accepted that theoretical terms have ‘excess content’ over the empirical manifestations with which they are associated. Correspondence rules do not define theoretical terms. To be sure, they do contribute to their meaning, by showing how these terms get connected to (and therefore get applied to) experience. But they do not exhaust the meaning of theoretical terms. But Carnap was unwilling to grant that theoretical terms are meant to designate unobservable entities. He summarised his new empiricist position as follows: “We can give [to questions of the sort ‘Does the electromagnetic field exist?’] a good scientific meaning, eg, if we agree to understand the acceptance of the reality, say, of the electromagnetic field in the classical sense as the acceptance of a language L_T and in it a term, say ‘E’, and a set of postulates T which include the classical laws of the electromagnetic field (say, the Maxwell equations) as postulates for ‘E’. For an observer to ‘accept’ the postulates of T, means here not simply to take T as an uninterpreted calculus, but to use T together with specified rules of correspondence C for guiding his expectations by deriving predictions about future observable events from observed events with the help of T and C” (1956, 45).

So Carnap wants to dissociate empiricism from a purely instrumentalist account of scientific theories, where theories are considered as merely syntactic constructs for the organisation of experience, for connecting empirical laws and observations that would otherwise be taken to be irrelevant to one another, and for guiding further experimental investigation. But he’s equally unwilling to accept that, say, the term ‘electromagnetic field’ refers to a n unobservable entity. He suggests that questions of reality of theoretical entities should give way to questions of language-choice, where a cartain theoretical language is chosen—and within it a certain theoretical term, e.g., ‘electromagnetic field’, in order to adequately describe observational phenomena.

So why did Carnap advance? He did do as a way to show how empiricists can meaningfully use tirreducible theoretical terms without thereby being committed to unobservable entities. But the two-tier model was meant to address some other issues, too, that had made other empiricists to move away from some empiricist tents. Carnap wanted to defend a sort of **semantic atomism** against Hempel’s (and Quine’s) **semantic holism**. He also wanted ot show that <u>not all</u> terms that feature in a theoretical vocabulary, or are tacked on to it, are meaningful. (If meaning holism is right then the latter seemed to be unavoidable.) Hempel, for instance, had suggested that theoretical terms get their meaning as a whole from the role they play within a theory. (“The cognitive meaning of a statement in an empiricist language is reflected in the totality of its logical relationships to all other statements in that language and not to the observation sentences alone” and “(M)eaning can be attributed only to the set of all the non-observational terms functioning in a given theory”). Carnap strongly disagreed with this. He thought that there was still space for a) defending the atomistic significance of theoretical terms, and b) drawing a boundary between meaningful and meaningless (cf. 1956, pp39-40). So, roughly, on the two-tier model, a theoretical term is meaningful not just in case it’s part of a theory. Rather, it’s meaningful iff (roughly) it makes some <u>positive contribution</u> to the experiential output of the theory. In other words, a term is meaningful iff its appearance in a theory makes some empirical difference. This move, Carnap thought, could render ‘ψ-function’ meaningful, although utterly remote from experience, but it wouldn’t thereby make meaningful any metaphysical speculation that is just tacked on to a scientific theory (cf. 1956, p39). There are some important technical problems with Carnap’s formal explication of this requirement. We shall not discuss them here, though. I will only say that Carnap’s account is too restrictive: it excludes too much. For it does not seem to be necessary for all meaningful theoretical terms to have straightforward experiential import, in the sense of yielding new observable consequences. Some theoretical terms may be introduced only in order to establish connections between other theoretical terms (e.g., the concept of ‘asymptotic freedom’ in quantum chromodynamics). It would be unwise to render such terms meaningless on the grounds that either alone or in conjunction with other terms they do not yield extra observable consequences. Conversely, Carnap’s theory is too permissive: for one can make *any* statement have some observational consequences by conjoining it with another that already has some. (If T entails O, then so does A&T, for any A. Rephrase A&T as T’ and you now have a new statement that entails observational consequences, and hence is meaningful on Carnap’s account.)

6.4 Semantic Realism

The crucial objection to the Carnap two-tier model, however, is this. Once it is accepted that T-terms have ‘excess content’ and once the old verificationism is abandoned, then it is but a short step to accept that T-terms have factual reference: they designate theoretical/unobservable entities. The so-called ‘surplus meaning’ of T-terms is grounded in their factual reference. As Feigl (1950, pp. 49-50) points out, in its treatment of the meaning of theoretical terms, verificationism runs together two separate issues: their “epistemic reduction (i.e., the evidential basis)” and “the semantical relation of designation (i.e., reference)” (1950, 48). Verificationism tends to conflate the issue of what *constitutes evidence for the truth* of an assertion with the issue of *what would make this assertion true*. (Grover Maxwell called the confusion of *what a statement asserts* with *how it comes to be known* the *fallacy of epistemologism*.) But if evidential basis and reference are separated, then the question of the meaning of theoretical terms is answered once and for all. Regardless of whether one acknowledges a difference with respect to *testability* between observational and theoretical assertions (a difference in pragmatics, not in semantics), both kinds of assertion should be treated semantically on a par, that is as being truth-valued. This simply requires that theoretical terms no less than observational ones have putative factual reference.

Following a suggestion made by Carnap himself in his (1939)—who followed Tarski’s work on truth—Feigl points out that the semantics of T-terms should pose no serious conceptual problem for empiricists. The designation of T-terms is specified in a rich enough meta-language according to rules of the form “'The magnetic field of the earth' designates the magnetic field of the earth”. Nothing more than that is required for semantics. This is clearly not enough to show that certain T-terms *have* factual reference, i.e., which entities we might be realists about. But this issue calls for a criterion that separates the terms that refer from those that do not. Feigl’s suggestion is this: “in the language of empirical science all those terms (and only those terms) have factual reference which are linked to each other and to the evidential basis by nomological relationships” (1950, p.50).

This suggestion may not sound enough to guide rational belief in certain existential claims. But rational belief in the reality of an entity ought to be guided by the confirmation that hypotheses about this entity enjoy. According to verificationism, something is deemed real if it is directly experienceable. But this restricted sense of reality goes with verificationism. Feigl points out that there is also an empirical sense of reality which is fully consistent with the new post-verification empiricism: something is real if it is confirmably required as having a place in the coherent spatio-temporal-causal account which science offers. On this account of ‘reality’ there is nothing to stop us from considering some putative referents to be real, insofar as, of course, the theories in which they feature are well-confirmed.

At any rate, empiricists, pretty much like anyone else, should leave it up to science to tell us what the world is like. What kinds of things exist in the world is clearly an empirical issue. What philosophers should do is analyse what we mean when we say that science describes a world where the *knowable* is in excess of the *known*. The suggestion that T-terms have factual reference aims precisely to address this issue: that there is more in the world than whatever falls within the reach of “the physically possible direct evidence”. A “full and just explication” of the way the language of science is used cannot do without designata of theoretical terms. For, short of full definability in terms of observables, what else could make theoretical assertions truth-valued? To be sure, the putative designata do fall outside the scope of direct experience. But theoretical assertions can enjoy confirmation by evidence. Yet, if theoretical assertions are not truth-valued, how can they possibly be confirmed or disconfirmed by evidence?

The position just outlined is the essence of what Feigl has called “semantic realism”. By and large, it forms the current common ground between scientific realism and modern empiricism. What modern empiricism challenges is not the semantic realist framework for understanding and deploying scientific theories. Instead, it challenges the ability of scientific method to produce a well-confirmed account of the unobservable structure of the world, and hence it challenges the rationality of belief in such an account. But we shall come backto this issue later on.

Although there is more to be said about how Carnap reacted to this challnge, all I want to stress here is the following: Carnap's travails made it plausible that theoretical terms are ineliminable and meaningful, in that a) their meaning is not exhausted by their conditions of application to experience and b) assertions employing them are not reducible to any set of observation sentences, no matter how big and detailed. Feigl's addition is that these finding can only make semantic realism a more plausible account of the semantics of scientific theories.

7. Observational-Theoretical Terms Distinction

As we have seen, already in 1936 Carnap argued there is no sharp dichotomy between observable and non-observable terms and predicates. However, in all of his relevant writings, he took the notion of being observable as relatively unproblematic and generally understood. Normally, he just offered examples of observable predicates such as 'red', 'blue, 'warmer than', 'large' etc. The whole programme of offering partial empirical meaning to theoretical term was developed as if there was such a dichotomy and unproblematic cases of observable predicates. Why did Carnap insist on this? I think Carnap perceived that such a distinction is necessary for the possibility of confirmation of scientific theories. Here is how the scheme is supposed to work. How are theories confirmed? By checking a subclass of their consequences. Which subclass? Those that can be overtly and intersubjectively checked as to whether they are true or false. The observational vocabulary is supposed to be able to express those consequences. Then the community can reach an agreement concerning whether or not the theory is confirmed. Besides, if theories issue in predictions couched in a common observational vocabulary, theories can be compared vis-à-vis their confirmation. If a theory T entails a prediction O and another theory T^* entails a prediction not-O, and if O obtains, then T is confirmed whereas T^* is disconfirmed. (Of course, I oversimplify the situation since I ignore the Duhem problem. But you get the idea.)

Be that as it may, the revolt against logical empiricism in the early sixties took as one of its most important tasks to uproot the alleged dichotomy between theoretical and observational terms. One main objection has been that although there are differences between observable and unobservable entities, they are so much diffuse and context-dependent that it is impossible to create a principled and absolute distinction between terms which refer only to unobservable entities and terms which refer only to observable ones. Take, for instance, the term ‘electron’. The term is theoretical when it is used in connection with the detection of particles in a cloud-chamber. But it becomes observational when it is used in connection with modern accelerators detecting the generation of other elementary particles. Based on such considerations, and lots of examples, Achinstein and others (e.g., Mary Hesse) suggested that, depending on the context, the set of theoretical terms is circumscribed differently (1965, pp.237-238). Given that some (probably most) terms which are classified as ‘theoretical’ in context C_1 may be classified as ‘observational’ in context C_2 (and conversely), it follows that there is no principled way to make sense of such a distinction, that is, there is no way in which we can separate out two classes of terms one comprising all and only observational predicates, the other comprising all and only non-observational ones.

Putnam pushed this line to its extremes, by arguing that “if an ‘observation term’ is a term which can, in principle, only be used to refer to observable things, then there are no observation terms (1962, p.218). For there is no single observational term which could not be used in application to unobservables without changing its meaning. (Putnam’s example is Newton’s reference to red corpuscles to explain red light. But other examples can be thought of.) And conversely, many theoretical terms can be seen as applying to observational states-of-affairs (e.g., when one says that the guy who stuck his finger in the mains was electrocuted.)

At any rate, Putnam argued, “the problem for which this dichotomy was invented (‘how is it possible to interpret theoretical terms?’) does not exist” (1962, p216). The dichotomy between observational terms and theoretical ones was devised in order to provide meanings to theoretical terms in virtue of an antecedently understood observational vocabulary. But theoretical terms can be fully understood, too. Their meaning is given by the theories in which they are embedded and is learned the same way in which the meaning of any term is, that is by learning a language, the language of theories in which these terms feature.

Another move worth noting is that adopted by Grover Maxwell. He suggested that observability is a vague notion and that, in essence, all entities are observable under suitable circumstances. ‘Observability’ should be best understood as being broadly synonymous with ‘delectability by means of some or other instrument’. If obsevability is understood thus, then “there is, in principle, a continuous series beginning with looking through a vacuum and containing these as members: looking through a window-pane, looking through glasses, looking through binoculars, looking through a low-power microscope, looking through a high microscope, etc. in the order given.” (1962, p.7). If, therefore, all there is a continuous degrees of observability, then then there is no natural and non-arbitrary way to draw a line between observational terms and theoretical ones: all terms apply to things that are to some degree observable. In fact, Maxwell employed this 'no-cut' argument in order to suggest something stronger than that, viz., that since 'theoretical' entities are ultimately observable, then we should not have any qualms about their existence, insofar as we don't doubt the existence of other observable entities. For if observability is a continuum, and if 'theoretical' entities are thus rendered observable to some degree and under certain circumstances, then it follows that "any talk of continuity from full-blown existence [direct observability] to nonexistence [indirect observability] is, clearly, nonsense" (op.cit., p. 9) For Maxwell the observable/nonobservable distinction "has no ontological significance whatever" (op.cit., p15). In other words, whether an entity is observable or not is nothing to do with whether or not this entity exists. The distinction is too anthropomorphic to have any ontological significance. Although this is sound, as far as it goes, van Fraassen has recetnly argued that this distinction might nonetheless has an epistemic significance: it is related to what we are in a position to believe in. But we shall come to this issue later on.

What is the point of the criticism of the distinction between obsevrational terms and theoretical ones. One should be careful here. The point of the criticism so far is not that there are no observational terms and predicates. Rather, the point is that any distinction between observational and theoretical terms is largely based on pragmatic considerations and therefore it has neither semantic nor epistemic significance. In a given context, or with respect to a given class of theories, one can distinguish between terms that are 'observational' (in the sense that the presence or absence of the entities they refer to can be easily agreed on, relying only on unaided senses, elementary instruments and commonly accepted background theories) and terms that are 'theoretical' (in the sense that the entities they refer to can only be indirectly detected or inferred). But such a distinction is neither absolute nor sharp. It's only purpose is to faciliate confirmation. One can now easily see that this way to challenge the alleged dichotomy between observational terms and theoretical ones does not block the possibility of confirming scientific theories. This rough pragmatic distinction of degree is all we need in order to confirm scientific theories. For scientific theories can be still seen as entailing predictions couched in a common 'observational' language (in the above broader and relativised sense), although the boundaries of this language shift and although some of the terms of this language would count as theoretical in a different context.

7. Theory-Ladenness of Observation.

There is, however, a much stronger way to read the claim that there is no sharp distinction between theoretical terms and observational ones. This is the view that, strictly speaking, there can be no observational terms at all. This view is inspired by the claim that all observation is theory-laden and concludes that there cannot possibly be a theory-neutral observation language. This is the view associated with Kuhn and Feyerabend.

7.1 Duhem's Thesis

The view that all observation is theory-laden goes back to Duhem. He suggested that "An experiment in physics is not simply an observation of a phenomenon. It is besides, the theoretical interpretation of this phenomenon" (1906, p.144). A typical case here is this: when a physicist sees, say, the pointer of an ammeter attached on a wire to move, she will normally describe her observation not by saying 'the pointer of the ammeter moves' but rather by saying that 'electric current flows through the wire'. As Duhem observed using a similar example, if she were asked to describe what she was doing, she wouldn't say that she was studying the movement of the pointer. Rather she would say she's measuring the intensity of the current. Observation in science is not just the act of reporting a phenomenon (whatever that means!). It is the interpretation of a phenomenon in the light of some theory and other background knowledge. In fact, strictly speaking, a phenomenon is an already interpreted regularity (or event).

More recent results in experimental psychology have shown that the theory-ladenness of observation extends to ordinary everyday cases. In the duck-rabbit case, for instance, one does not merely observe a shape composed of certain curved lines. One sees a rabbit or a duck or both. The perceptual experience is theoretically interpreted. This interpretation is, for the most part, unconscious. But, as you can clearly see, there are different possible theoretical interpretations of the same perceptual experience. (For more on this one can see Richard Gregory's book The Intelligent Eye, and Gillies's Philosophy of Science in the Twentieth Century.)

Kuhn and Feyerabend pushed the theory-ladenness-of-observation thesis to its extremes, by arguing that each theory (or paradigm) creates its own experiences (in fact it determines the meaning of all terms—be they 'observational' or 'theoretical'—occurring in it) and that there is no neutral language which can be used to compare different theories (or paradigms). For Kuhn two paradigms are incommensurable just in case there is no “language into which at least the empirical consequence of both can be translated without loss or change”. As Feyerabend once put it: “the meaning of observation sentences is determined by the theories with which they are connected. Theories are meaningful independent of observations; observation statements are not meaningful unless they have been connected with theories. (...) It is therefore the observation sentence that is in need of interpretation and not the theory”.

We have seen already that one major argument for the meaningfulness of theoretical terms is that they get their meaning from the theories and network of laws in which they are embedded. This view has been called semantic holism. Advocated by Quine, Hempel and later on by Kuhn and Feyerabend, it has significantly contributed to the wide acceptance of the claim that theoretical discourse is meaningful. Now we see, however, that combined with the thesis that all observation is theory-laden, it can lead to the conclusion that the meaning of observational terms is also specified in a holistic way. Worse than that, since the meanings of terms is determined by the theory as a whole, it can be now claimed that every time the theory changes, the meanings of all terms change, too. This is the so-called radical meaning variance. But then, there is simply no way to compare theories. This is, in broad outline, the genesis of the thesis of semantic incommensurability. We can now easily see that if this thesis is right, then not only is the realist project of finding true theories of theworld in danger, but also the more modest empiricist project of finding emprically adequate theories is in danger too. Both the realist and the empiricist projects require some meaning invariance in order to get off the ground and be able to talk about theoretical and/or empirical progress.

Before we discuss incommensurability in some more detail, it is instructive to see what Duhem (and I, on his behalf I hope) think his thesis entails. Duhem notes that “what the physicist states as the result of an experiment is not the recital of observed facts but the interpretation and transposing of these facts into the ideal, abstract, symbolic world by the theories he regards as established” (1906, p.159). This entails that without knowing these theories, we cannot “understand the meaning he gives to his own statements”. So far so good. But does this generate insurmountable problems in our attempt to evaluate the physicist’s work, understand his results, see whether they confirm the theory etc.? Not really. First of all, it may be the case that this physicist’s theories “are those we accept”. Then, given that we follow the same rules in the interpretation of the same phenomena, we can understand what he’s doing. Second, even if he advocates a different theory to ours, there are two options open to us: we can either try to determine the extend to which the phenomenon can be interpreted in the light of commonly accepted background theories. Or we can examine whether there is a common sub-language (not necessarily observational, even in the loose sense discussed above) between our theories, which can describe the phenomenon at hand. If either of these options fail, then we must “try to establish a correspondence between the theoretical ideas of the author we are studying and ours”. There is no reason why we shouldn’t succeed in doing so. Duhem in fact presents a few interesting example from the history of optics. This ‘correspondence’ need not involve a word-by-word translation between the two theories. Nor is it the case that, necessarily, the whole of one’s theory is involved in the interpretation of an experiment. If there is no way in which such correspondence can be established—not even a local one, that is—then we cannot possibly make sense of our physicist’s experiment. Early Kuhn (and Feyerabend) attempted to show that this is what typically is the case. But as we shall see this ain’t necessarily so.

7.2 Incommensurability

We’ve already explained in the last chapter how Kuhnian paradigms are said to be incommensurable. We can summarise it thus: the two paradigms are incomparable. The new paradigm a) defines a new set of problems, and sets new standards as to what constitute legitimate solutions to those problems; b) the meaning of concepts (intension as well as extension) change radically; c) a different phenomenal world is ‘created’.

Let me focus on semantic incommesurability. If the paradigm as a whole determines the meaning of all terms and predicates, then paradigm change implies meaning change. However, meaning variance, even radical variance, is consistent with sameness in reference. Admitting that there are changes in meanings when the relevant paradigms change is compatible with the claim that all these conceptual shifts are about the same entities in the world. For instance, although the meaning of the term ‘whale’ has changed since the term was accommodated in the modern framework of biological sciences—whale is a mammal—the referent of ‘whale’ is still whales. In other words, the meaning of terms may depend on—and change in accordance to—the “theoretical context” (Feyerabend) or the paradigm (Kuhn) in which they occur and yet their referents may remain the same. However, Kuhn insists that it is reference variance that occurs just as well when the meaning of a term changes. One of the standard examples of reference change is that of the terms ‘mass’, ‘space’ and ‘time’ as they occur in Newton’s theory and in Einstein’s theory. As Kuhn stated “the physical referents of these Einsteinian concepts [i.e., mass, space and time] are by no means identical with those of the Newtonian concepts that bear the same name” (1970, 102). But Kuhn also suggested that terms such as ‘planet’ refer to different things when they occur in different theories (e.g., in Ptolemy’s theory ‘planet’ refers, among other things, to the sun but not to the earth, while in Copernicus’ theory, ‘planet’ refers to the earth but not to the sun) (1970, 128). If Kuhn is right, semantic incommensurability follows straightaway. For then there are no common semantic resources between different paradigms. Both the sense and the reference of a word change and the use of the same word in two different paradigms is a mere equivocation. Take an example. In Newton's theory mass is an invariant property. In Einstein's theory mass varies with velocity. On the face of it, these two claims are incompatible. One is inclined to think that they make two incompatible claims about the same entity and at least one of them should be false. For Kuhn, however, things are different. The impression of incompatibility arises through the use of the same term. But this is an equivocation. To remove it, we must use different terms. The statement 'mass$_1$ is an invariant property' and 'mass$_2$ depends on velocity' are not incompatible. The two terms 'mass$_1$' and 'mass$_2$' have different meanings as well as different reference. But then there is no common ground for semantic assessment of the two statements above.[3]

Kuhn's view has been challenged on many grounds. First of all, it entails the rather absurd conclusion that the meaning and reference of a term changes whenever there occurs the slightest change in the network theories in which it is embedded. One can argue here that, after all, semantic holism can be moderate. It may suggest that terms do not get their meaning in isolation but within a network of laws and theories in which they are embedded and yet also urge that not all parts of the network are equally responsible (or inextricably interlocked) in fixing the meaning of terms; a change in one term may not affect all the rest within a given network. In fact, Feyerabend accepted that not all theoretical changes lead to changes in meaning and reference. He suggested that the rules (assumptions, postulates) of a theory form a hierarchy where more fundamental rules are presupposed by less fundamental ones. Then, only changes in fundamental rules lead to changes in meaning and (possibly) reference. Second, even if one accepts that there are meaning changes whenever there are radical changes in theory, one can resist the conclusion that reference changes, too. For instance, one may follow the so-called 'causal theories of reference', due to Kripke and Putnam, in order to argue that salient scientific terms, such as 'electricity', 'heat', 'electron' etc, are trans-theoretical because they refer to the same things, viz., to the causes of the phenomena that these terms were introduced to describe (cf. Putnam, 1975, p202). This theory disposes of (semantic) incommensurability, since all different theories of, say, electricity, talk about and dispute over the same thing, viz. electricity, or better the causal

[3] After 1969 Kuhn restricted the notion of incommensurability to solely semantic incommensurability. There is still meaning variance but it's local rather than global and it concerns "only small groups of terms". After 1980, he suggested that incommensurability amounts to untranslatability in the technical sense of the lack of a systematic replacement of words or groups of words by ones equivalent in meaning. Theories are incommensurable, if untranslatable in this sense. But, theories may be interpretable (in a loose sense) into one another. (For more details on the development of Kuhn's notion of incommensurability, you can look at Hoyningen-Huene's book, section 6.3).

agent of electrical effects. Because of its relevance to the realist project in general, I shall discuss the causal theory in some detail.

8. The Causal Theory of Reference

According to the received description theories of reference, the reference (or denotation) of a referring expression (e.g., a proper name, or a singular term) is specified by means of a description (normally understood as specifying the sense of the referring expression). So, on this theory, each term (or proper name) is associated with either a unique propositional (attributive) description or, in a more sophisticated manner, with a cluster of (probably weighted) descriptions. The unique individual picked out by this description, or cluster thereof, is the referent of the term. If the description, or a weighted most of the descriptions associated with a term t is satisfied by an individual y, then y is the referent of t; but if nothing satisfies the description, or a weighted most of the descriptions, t does not refer. The thrust, as it were, of the description theory is that the relation between a word and its referent is mediated by the sense of the word (a concept). So, an expression acquires its reference (if any) via its sense. The main problem with the descriptive theories is that they, generally, associate too rich a description with a term/name. It is not necessary, sometimes not even true, that the individual referred to satisfy all (even most) of the descriptions associated with the name.[4]

It is worth stressing, in passing, that the traditional description theory of reference is not, necessarily, holistic. It allows for weighted descriptions of individuals. Then, not any and every change in the cluster of descriptions will yield reference variance. Besides, it allows that two different descriptions pick out the same individual (i.e., that they may be coreferential) provided that they are not inconsistent. Hence, it's not the description theory on its own that generates incommensurability. Rather, it is only in conjunction with a radical holistic theory of meaning that it might.

The Causal theory of reference was, initially, introduced by Kripke as a way to avoid the shortcomings of the description theory. Kripke identifies reference with a causal—historical chain which links the current use of a term with an act of baptism, where a name was picked to dub an individual. Descriptions associated with the name might (all) be false, and yet the user of the name still refers to the individual dubbed, insofar as his (the user's) use of the name is part of a causal transmission chain that goes back

[4] The description theory, in both of its classical (Frege-Russell) and modern (Wittgenstein-Searle) forms has been criticised and showed inadequate by Kripke (1972); cf. also Devitt & Sterelny (1987, chapter 3).

to the dubbing ceremony. Occasionally, the introducing event may involve *some* description of the entity introduced. In fact there are cases in which the introduction is made *only* via a description, e.g., the introduction of the planet Neptune or of the 'Jack the Ripper' (cf. Kripke, 1972, 79-80 & 96). But the description is not analytically tied with the term. It rather "fixes the reference by some contingent marks of the object" (op.cit., 106). Generally however, what fixes the reference of a term is not the descriptions associated with it but the causal chain which connects the term with the object named. So, the thrust, as it were, of the causal theory of names is that the relation between a word and an object is direct, unmediated by a concept. Causal theories dispense with senses as reference-fixing devices and suggest that the reference of a word is whatever entity "grounded" the word in a certain "dubbing ceremony" in which the word was introduced.

Putnam, more than anybody else, saw that this theory can be easily extended to 'natural kind' terms as well as to 'physical magnitude' terms. The reference of a natural kind term (or, for that matter of a theoretical—physical magnitude—term) is fixed during an introducing event, i.e., an event during which the term is given its referent. According to Putnam, reference is fixed by "things which are given existentially" (1983b, 73). In the case of a natural kind term, this means that one picks out by ostension an object (e.g., a tiger), attaches a name to it and asserts that this name applies to all and only the objects that are similar to the one picked. (Similarity need not concern the manifest properties. If it is to describe natural kinds, then it's got to relate to underlying properties. For instance, 'water' does not refer to the stuff that has this and that manifest property, e.g., transparent, odourless etc., but to the stuff that is H_2O.) So, "a term refers (to the object named) if it stands in the right relation (causal continuity in the case of proper names; sameness of 'nature' in the case of kinds terms) to these existentially given things" (Putnam, 1983b, 73). When the introductory event is completed, the term is transmitted through a linguistic community. The term is borrowed by other users, this borrowing being reference-preserving if the users are connected to the introductory event with some causal chain of term-transmission.

When it comes to the reference of 'physical magnitude' terms, the causal theory suggests the following: when confronted with some phenomena, it is reasonable to assume that there is something, i.e., a physical magnitude or process, which causes them. Then we (or indeed, the first person to notice them) dub this magnitude with a term t and associate it with the production of these phenomena (e.g., the sensation of heat—cf. Kripke, 1972, 132-133 & 136). This is the introducing event of t as referring to this magnitude. Here again, one may associate the term with a description, i.e., with a causal story, of what the nature of this magnitude is and in virtue of what it causes the observable effects. However, one's initial description may be incomplete and even misguided. It may even be a wrong description, a totally mistaken account of the nature of this causal agent. Nonetheless, one has introduced existentially a referent—an entity causally responsible for certain effects. One has asserted that:

> Bα {α is causally responsible for certain phenomena F and (Ax) (x is an α if and only if x picks out the causal agent of F)}.

The intuitive appeal of the causal theory of reference rest on the following difference: it is one thing to assert that there is an entity to which t refers; it is quite another matter to find out the exact nature of this entity (i.e., the referent α of t) and hence the correct description associated with t. Our beliefs about this entity may be incorrect. They may change as our knowledge of its nature advances. But our initial positing of an entity causally responsible for these effects will not change (cf. Kripke, 1972, 138-139; Putnam, 1975a, 197 & 204).

One can see that the causal theory disposes of (semantic) incommensurability, since all different theories, say of electricity, talk about, and dispute over, the same thing, viz. electricity; or better the causal agent of electrical effect. It allows that even though past scientists had partially or fully incorrect beliefs about the properties of a causal agent, their investigations were continuous with those of subsequent scientists since their common aim has been to identify the properties of the same entity, i.e., of the causal agent of certain phenomena.

The causal theory also yields that the determination of the reference (and of the meaning) of a term is, by and large, an issue which cannot be solved a priori by means of conceptual analysis, but rather with empirical investigation into the features of the world and the natural kinds occurring in it (cf. Kripke, 1972, 138). How the world is is an indispensable constraints on the theory and practice of fixing the reference (and meaning) of the language we use to talk about it.

However, the causal theories exhibit some important shortcomings which have been discussed extensively in the relevant literature (cf. Fine, 1975; Papineau, 1979, 161 & 165; Sterelny & Devitt, 1987, 72-75). Their essence is that, generally, reference-fixing cannot be as description-blind as the causal account suggests. Let me just concentrate on the reference of theoretical terms and explain some typical shortcomings of the causal theories.

As we saw above, the causal theory suggests that reference is fixed by things which are given purely existentially. But, when it comes to the reference of theoretical terms, ostention cannot be of any help. When, for instance, Benjamin Franklin introduced the term ‘electricity’ what he offered was something like this: there is a physical magnitude which causes sparks and lightning bolts—adding, possibly that electricity is capable of flow or motion (cf. Putnam, 1975, 199). That is, the magnitude which ‘electricity’ was coined to refer to was given by stating some of its manifest effects and, possibly, an elementary description, i.e., that, whatever else it is, it is capable to flow. ‘Electricity’ could have been introduced on a different occasion. In fact, André Ampère also introduced ‘électricité’ by means of a different description of the effects it produces, viz. currents and electromagnets. What is there in common in all occasions where ‘electricity’ was introduced or could have been introduced? What is there in common between Franklin’s ‘electricity’, Ampère’s ‘électricité’ and indeed, anybody else’s ‘electricity’? Putnam response is that “that each of [the occurrences of the term ‘electricity’] is connected by a certain kind of causal chain to a situation in which a description of electricity is given, and generally a causal description—that is, one which singles out electricity as the physical magnitude responsible for certain effects in certain ways (1975, 200).

Given, however, that all these descriptions may have been misdescriptions of electricity (rather than descriptions of nothing at all) (cf. Putnam, op.cit., 201), it seems that the foregoing response amounts to the following: what there is in common in all occurrences of the term ‘electricity’ in different theories and descriptions is electricity itself, i.e., the physical magnitude causally responsible for electrical phenomena. This physical magnitude is the referent of the term ‘electricity’ and guarantees the sameness in reference of the occurrences of this term. But then, in the case of the reference of theoretical terms, the ‘existentially given thing’ is nothing but a causal agent, an agent with the causal capacity to produce certain effects.[5] A quick worry here might be that there is no guarantee that there is just one causal agent that causally produces behind all these phenomena, e.g., electric currents, lightning bolts, deflections of magnets etc. This might well be so, but the causal theorist will immediately add that he is not concerned with the epistemological problem of how we are able to assert that all these phenomena are due to electricity. All he’s concerned with is showing how all these different terms may nonetheless refer to the same entity. If it happens that electricity is not responsible for, say, lightning bolts, then Franklin’s ‘electricity’ does not refer to electricity.

[5] Occasionally, Putnam seems to suggest that “an approximately correct definite description” is required for successful reference (cf. 1975, 200).

A more promising critique is to say that given that causal theory reduces referential stability to the bare assertion that a causally efficacious agent exists behind a set of phenomena., continuity and sameness in reference become very easily satisfiable. If the causal agent of some phenomena is given only existentially, and if any description of its nature associated with the relevant term may well be false, then the term will never fail to refer to something: to whatever causes the relevant phenomena—provided of course that these phenomena do have a cause. To put the same point negatively, it is not clear at all what could possibly show that the entity posited does not exist; that is, it is not clear under what circumstances there is referential failure.

Take, for instance, the case of phlogiston. ‘Phlogiston’ was introduced on many occasions by means of a causal description, i.e., one that singled phlogiston out as the physical magnitude causally involved (given off) in combustion. Phlogiston, however, does not exist, nor is it causally involved in combustion. Instead, we now think that oxygen is. Does this mean that phlogiston theorists had been referring to oxygen all along? If the reference of ‘phlogiston’ was coined to refer purely existentially to whatever is causally involved in combustion, then it seems inescapable that ‘phlogiston’ refers to oxygen, after all. Then, there is little space for referential discontinuity between theories. Stability comes too cheap. In order to say (the correct thing) that ‘phlogiston’ refers to nothing, we need to say that there is nothing in nature that possesses the properties that phlogiston was supposed to possess in order to play its assigned causal role in combustion. That is, we need to say that there is nothing that satisfies the description ...phlogiston.... More generally, there is no way of telling of a putative causal agent that it does not exist apart from showing that there is no entity possessing the properties attributed to this agent. This procedure involves examining whether the descriptions associated with the term that purports to refer to this agent are satisfied. Referential failure cannot be assessed without appealing to some description, and I content, by symmetry, so is the case for referential success. Causal theories of reference are certainly right to emphasise that reference involves a causal process which connects the term with its referent. But reference-fixing can be a purely causal process only at the price of making referential stability too easy (and too trivial) and referential failure rather impossible.

That some descriptions about what sort of entity is the referent of a term, or about what major characteristics it has, are necessary in reference-fixing becomes apparent when we want to judge whether two distinct terms might refer to the same entity. Defending this possibility is central to a realist account of science. In order for realists to defend that there is some substantive continuity in revolutionary theory-change, they have to show that distinct terms in different theories can refer to the same entity in the world, although they characterise it differently. If past theories are to be claimed to be approximately true, then it should be at least the case that they referred to the same entities as successor theories do and made some true (or approximately true) assertions about them. (We'll come back to this is issue when we discuss the so-called argument from the pessimistic induction against scientific realism.)

If realists adopt a pure causal theory of reference, then the above problem seems easy to solve. Take two distinct theories of light. Fresnel's theory introduces the term 'luminiferous ether' to refer to the material medium through which light-waves propagate, a medium consisting of ethereal molecules , having solid-like properties so that it can sustain transverse waves. Maxwell's theory—in its mature form—dispenses with a material medium for light-propagation and instead introduces a 'disembodied' medium (in effect, the physical space) to which the term 'electromagnetic field' refers. However, as Hardin and Rosenberg (1982, 613) note, the luminiferous ether has played "causal role we now ascribe to the electromagnetic field". On their view, if one allows that reference follow causal role, and given that ether and electromagnetic field played the same causal role with respect to optical and electromagnetic phenomena, it seems not unreasonable "to say that 'ether' referred to the electromagnetic field all along" (op.cit., 614).

One should tread carefully here. This variant of the causal theory does indeed show how to distinct terms can refer to the same entity, in fact to an entity we now posit. But at a price. As Laudan put it, the 'sameness-of-causal-role' account of reference "confuses a shared explanatory agenda (i.e., common problems to be solved) with a shared explanatory ontology (i.e., the characteristics of the postulated explanatory entities)" (19842, 161). Clearly, a mere similarity in the phenomena to be accounted for does not warrant sameness in the underlying structures that cause these phenomena. After all, one may think, it may be the case that the underlying putative causes are merely analogous, but not identical. Laudan goes on to stress that "to make reference parasitic on what is being explained rather than on what is doing the explaining entails that we can establish what a theory refers to independently of any detailed analysis of what the theory asserts" (ibid.).

Laudan's objection is fair. He is right to stress that what the advocates of referential stability in theory-change must show is not just continuity in the phenomena to be explained. Important though that is, it is not sufficient for continuity at the level of entities that are posited to explain these phenomena. That is why in judging continuity of reference we need to appeal to some descriptions. For how else can we explain —in a non-trivial way—that term t and term t' have the same reference, except by showing that at least some of the descriptions of what t refers to are also descriptions of what t' refers to? Then, judgements of successful reference to physical magnitudes, causally responsible for some observable phenomena, rest on taking on board, at least some, descriptions about the causal role of this magnitude.

If some descriptions are involved in reference-fixing, then we are pushed back towards description theories in order to account for their use in reference-fixing. I take it then that a correct theory of reference must utilise resources from both causal and descriptive theories of reference. I shall refrain from developing such a theory here, but it can be shown to be available. My own view is that a realist theory of referential stability can and should attempt to include some substantive continuity at the level of properties that the putative referents are thought to possess, properties in virtue of which they play the causal role they are ascribed. Such an account of reference will then be able to ground and explain the claim that the posited entities share the same causal role, thereby meeting Laudan's objection. The bare essentials of a rather satisfactory account are this: the reference of a theoretical term is indeed fixed during an introductory event in which a causal agent is posited to account for some phenomena. But what some descriptions about this entity is like are also involved. What kinds of descriptions? Description that state in general terms the properties and mechanism through which this causal agent brings about its effects. How detailed these descriptions are in a matter of debate. But that they should be there is beyond doubt. Then a term t refers to an entity x iff x is causally responsible for the phenomena for which x was introduced and x possesses the most fundamental properties in virtue of which it is supposed to causally produce these phenomena. (Since these properties are mentioned in the relevant description associated with t , we can simply say that x should satisfy these descriptions.) Similarly two distinct terms t and t' in theories T and T' respectively refer to the same entity x iff i) the posited entity x plays the same causal role in both theories vis-à-vis the same phenomena; and ii) the most fundamental properties in virtue of which the posited entity x plays its causal role according to T, are also taken to be fundamental properties of the posited entity x, according to T'. This account, however, is only a sketch that needs to be fully developed.

7. Concept Empiricism and Realism without Tears?

Suppose one wants to stick with concept empiricism and advocate the Principle of

Acquaintance: ‘In meaningful discourse every non-descriptive term is known by acquaintance’. Can one be any kind of realist about unobservable entities, or is one bound to believe that no meaningful assertions about unobservable entities can be made? Grover Maxwell suggested that it is perfectly possible for someone to be a concept empiricist and yet a realist about unobservables. He called his position structural realism. (This is not the same as Worrall’s structural realism. But it can be seen as providing the general framework for the development of Worrall’s position. More on Worrall’s position later.)

Maxwell (1970) suggested that if one accepts the Ramsey-sentence of a theory, then one can treat all theoretical terms as existentially quantified variables about which one can make meaningful assertions involving only observational terms (ie, known by acquaintance) and logical symbols. In order to get the Ramsey-sentence ^{R}T of a (finitely axiomatisable) theory T we proceed as follows. (This was first suggested by Frank Ramsey in his piece Theories in 1928.)

We write down the theory T in a first-order language, and partition its descriptive vocabulary in two sets, one containing theoretical terms the other containing observational ones. We then conjoin all the axioms of T into a composite axiom expressed by a single sentence S. Then, whenever a theoretical predicate, say Q, occurs in S we replace it by a second-order variable α. (That is a variable ranging over properties or sets of individuals.) If another theoretical predicate, say L, occurs in S, we replace all of its occurrences with another variable, say β. We repeat this procedure for all theoretical predicates of S, until we have replaced them all with variables. So, we end up with another formula S' which is like S except that it contains variables in place of the theoretical predicate of S. Finally, we bind these variables by placing an equal number of existential quantifiers $\exists\alpha$, $\exists\beta$, ... in front of S'. We have then constructed the Ramsey-sentence ^{R}T of T. (Example: Take the very simple theory “Water contains hydrogen atoms and oxygen atoms in proportion 2:1 and is tasteless, odourless and transparent”. The theoretical predicates here are ‘hydrogen atoms’ an ‘oxygen atoms’ while the rest are either logical terms (eg, proportion) or observational terms. We first replace all occurrences of theoretical predicates with variables and get the open formula: ‘Water contains α and β in proportion 2:1 and is tasteless, odourless and transparent’. Then we bind these variables with existential quantifiers and with a bit of rearranging and symbolic notation we get: $\exists\alpha\ \exists\beta\ \forall x$ (Wx then xCα and xCβ and β(2:1)α and Tx and Ox and Rx).)

The Ramsey-sentence of a theory is such that its only descriptive terms refer to observables. On the other hand, although the theoretical predicates are absent—and hence although, strictly speaking, the issue of their meaning is not raised—every unobservable property and class of entities referred to by the original theory is still referred to by the Ramsey-sentence. The Ramsey-sentence has provably the same empirical content with the original theory. But it says more than this empirical content. It asserts that there exist properties and classes of entities that stand in specific relations to one another and generate certain observable consequences. We may not know *what* these properties and classes are, but we know that they *are* (ie, that they exist). (A closer-to-home analogue of the Ramsey-sentence is the sentence ‘Someone stole my car last night’. In this sentence, if true, one refers to the thief without knowing him (her) and therefore without naming him (her).) In other words, the Ramsey-sentence of a theory, preserves the *functional role* that theoretical entities play within a theory, without being committed to what exactly these entities are—apart from the claim that they generate certain observable consequences and stand in certain relations to one another. It should then be clear that the Ramsey-sentence formulation of a theory does not make theoretical entities redundant. Nor does its truth-value relate only to the observable consequences of the theory.The construal of theoretical terms as existentially quantified variables yields a theoretical formulation which is either true or false, its truth-value depending on the existential claims about unobservables and their relations that the Ramsey-sentence makes.

Given that he Ramsey-sentence of the theory preserves the structure of the original theory, Maxwell suggests that the ‘Ramsey way’ is best understood as “structural realism”. As such, it suggests that i) scientific theories issue existential commitments to unobservable entities and ii) all non-observational knowledge of unobservables is *structural knowledge*, i.e., knowledge not of their first-order (or intrinsic) properties, but rather of their higher-order (or structural) properties (cf. 1970; 1970a). In Maxwell’s own words: “our knowledge of the theoretical is limited to its purely structural characteristics and (...) we are ignorant concerning its intrinsic nature” (1970a, p.188). Maxwell thought that this was essentially Russell’s insight in his *The Analysis of Matter*.

Despite that this view makes concept empiricism co-habit with realism, it faces some rather severe problems *qua* realist position. Here I will only raise briefly some objections.

Whether an interpreted theory is true or false is surely an empirical matter. But the Ramsey-sentence of a theory is *guaranteed* to be satified, i.e., it is guaranteed that there is an interpretation of the second-order variables which makes the Ramsey-sentence true. In order to show this, let’s take a simple example. Take the ‘theory’ $\forall x\,(Px \varnothing Ox)$, where ‘P’ is a theoretical predicate and ‘O’ an observational one. Write down its Ramsey-sentence: $\exists\varphi\forall x(\varphi x \varnothing Ox)$. It is easy to see that now matter how one interprets ‘O’, there *always* will be some interpretation of φ which makes the Ramsey-sentence true (other than the obvious candidate, i.e., the empty set). The Ramsey-sentence above is equivalent to $\exists\varphi\forall x(\square\varphi x \; \Delta \; Ox)$. All this asserts is the trivial claim that there is a property that either every x does not possess or every x possess O, no matter what ‘O’ is. (If all you say is that there is such a property, then here is one: Instantiate φ as ‘O’, no matter what ‘O’ means, and the Ramsey-sentence above turns into a logical truth.)[6] If, in response to this, the structuralist tries to impose restrictions to the interpretation of the range of φ, it is not clear at all that this can be done without abandoning the claim that only structure can be known.

This last point is merely an instance of a bigger problem that has been recently brought in focus by Demopoulos and Friedman (1985), although its original author is the mathematician M. H. A. Newman (1928): that structural realism is either trivially true or incoherent. Here is why.

The structural realist position is that all that is known about theoretical entities is a purely structural claim: that they stand in some complicated relation to the observable phenomena. Their intrinsic properties—what these entities are (Russell’s “qualities”)—are deemed irrelevant to the functioning of the theory and to the understanding of the world. Russell went as far as to say that “nothing in physical science ever depends on the actual qualities” (1927, p.227). And that “the only legitimate attitude about the physical world seems to be one of complete agnosticism as regards all but its mathematical properties” (op.cit., pp.270-271).

Newman’s critique of Russell’s position is, in essence, the following. In order to give some structure to a particular domain A, one must first specify the relation that structures this domain. Can the structuralist say of this relation only that it exists? No. One cannot merely say there exists a relation R that structures a particular domain A, but of the relation nothing is known but the structure it generates. This claim is trivial and, in fact, it offers no information about the domain other than its cardinal number. The reason is simply that provided that a domain A has enough individuals, it can possess any structure whatever compatible with the cardinality of the domain. No empirical investigation is necessary to find such a strucure. If, however, the structural realist tries to specify which relation is the appropriate to choose, he will have to go beyond his structuralism and talk about the nature of the relations. (This is the line that Russell took. The ‘important’ relation is causal continuity.)

From an intuitive point of view, Newman points to the following problem: the structure of a certain domain can be represented by means of a relational graph, where arrows connect

[6] This objection was first raised by Scheffler in his (1963, 218).

some dots iff the two individuals represented by the dots stand in the relation R represented by the arrow. Even if we specify the domain of discourse, and even if the relational structure is such that it gets connected with some observations, there still remains the issue of what exactly this relational structure is, i.e., what its intended interpretation is. Structural knowledge alone cannot possibly specify this: too many different theoretical relational structures can be defined on the same domain of individuals such that they account for exactly the same observations, i.e., they surface in the same way. In order to pick one of them, we certainly need to go beyond structure.

This is precisely what the scientific realist does. Generally, she would dispense with the Ramsey-sentence and go for the interpreted theory with its intended existential commitments. Or, insofar as she endorsed the Ramsey-sentence approach, she would appeal to the language of physics and its intended interpretation. In doing so, she would talk about the qualities of the theoretical entities posited by physics and would deem important those relations that the physical theories deem important. But she wouldn't thereby commit herself to a distinction between structure and quality. She didn't impose such a distinction in the first place. The structural realist, on the other hand, would either have to remain silent or to undermine the very distinction upon which she bases her epistemology and her understanding of theories.

Study Questions

1. Explain how reduction sentences work. Can they be used to define theoretical terms?

2. Can we draw a sharp distinction between observational and theoretical terms? Justify your answer.

3. State and discuss Duhem's thesis that all observation is theory-laden. Does Duhem's point necessarily lead to the incommensurability thesis?

4. What does the Kuhnian notion of incommensurability involve?

5. Critically examine Putnam's causal theory of reference. (Work on his piece 'Explanation and Reference'.) How does it solve the problem of semantic incommensurability?

6. Is the principle of acquaintance consistent witrh realism about unobservables?

References

Achinstein, P. (1965) 'The Problem of Theoretical Terms', American Philosophical Quarterly, **2**, No.3—reprinted in B.Brody (ed.) Readings in the Philosophy of Science, (1970), Englewood Cliffs, NJ: Prentice-Hall Inc.

Carnap, R. (1928) The Logical Structure of the World, (trans. R George), Berkeley: University of California Press.

Carnap, R. (1936) 'Testability and Meaning', Philosophy of Science, **3**, pp.419-471.

Carnap, R. (1937) 'Testability and Meaning —Continued', Philosophy of Science, **4**, pp.1-40.

Carnap, R. (1939) 'Foundations of Logic and Mathematics', International Encyclopaedia of Unified Science, **1**, No.3, Chicago IL: The University of Chicago Press.

Carnap, R. (1956) 'The Methodological Character of Theoretical Concepts' in H. Feigl & M. Scriven (eds.) The Foundations of Science and the Concepts of Psychology and Psychoanalysis, Minnesota Studies in the Philosophy of Science, **1**, Minneapolis: University of Minnesota Press.

Demopoulos, W. and Friedman M., (1985) 'The Concept of Structure in *The Analysis of Matter*', Philosophy of Science, reprinted in C. Wade Savage & A. Anderson (eds) Rereading Russell, Minnesota Studies in the Philosophy of Science, **12**, Minneapolis: University of Minnesota Press.

Feigl, H. (1950) 'Existential hypotheses: Realistic versus Phenomenalistic Interpretations', Philosophy of Science, **17**, pp.35-62.

Fine, A. (1975) 'How to Compare Theories: Reference and Change', Nous, **9**, pp.17-32.

Hempel, C. (1963) The Implications of Carnap's Work for the Philosophy of Science, in P.Schilpp (ed.) The Philosophy of Rudolf Carnap, La Salle, IL, Open Court.

Maxwell, G. (1962) 'The Ontological Status of Theoretical Entities', Minnesota Studies in the Philosophy of Science, **3**, Minneapolis: University of Minnesota Press.

Maxwell, G. (1970) 'Structural Realism and the Meaning of Theoretical Terms', in Analyses of Theories and Methos of Physics and Psychology, 4, Minneapolis: University of Minnesota Press.

Putnam, H. (1962) 'What Theories are Not', in Putnam's, Mathematics, Matter and Method, Philosophical Papers Vol.1, Cambridge: Cambridge University Press.

Putnam, H. (1975a) 'Explanation and Reference' in Philosophical Papers, Vol.2, Cambridge: Cambridge University Press.

III. Instrumentalism and Realism

9. Instrumentalism and Atoms

In his unsigned preface of Copernicus' On the Revolutions of the Celestial Spheres, (1543), **Andreas Osiander** stated:

"the) astronomer's job consists of the following: to gather together the history of the celestial movements by means of painstakingly and skilfully made observations, and then—since he cannot by any line of reasoning reach the true causes of these movements—to think up or construct whatever hypotheses he pleases such that, on their assumption, the self-same movements, past and future both, can be calculated by means of the principles of geometry. (...) It is not necessary that these hypotheses be true. They need not even be likely. This one thing suffices that the calculation to which they lead agree with the result of observation".

Although, Osiander talks only about astronomy, this is one of the most accurate statements of the instrumentalist conception of scientific theories. Theories are not supposed to offer a true description of the phenomena, but rather, to save the phenomena, that is to offer a (mostly mathematical) framework in which the phenomena can be embedded. On the contrary, the realist conception of scientific theories—towards which Copernicus himself was inclined—is that, as **Duhem** put it (for the case of astronomy), "a fully satisfactory astronomy can only be constructed on the basis of hypotheses that are true, that conform to the nature of things" (1908, p62).

Scientific theories, especially during the eighteenth and nineteenth centuries, posited a number of unobservable entities and processes, such as light corpuscles, light-waves, the 'luminiferous ether', molecules, atoms, various forces, etc. According to the realist conception, these posits are attempts to characterise "the nature of things", to find out about the 'furniture of the world', to explain the phenomena. But if scientific theories have only instrumental value, then what is the purpose of positing all these unobservable entities?

Instrumentalists have an answer that goes back to **Ernst Mach**. Under the motto 'science is economy of thought', he suggested that the aim of science is to classify appearances in a concise and systematic way; or as he put it: "to replace, or save, experiences, by the reproduction and anticipation of facts in thought" (1893, p577). So, for instance, instead of noting individual cases of light-refraction, we embody them in a single expression, the law of refraction (sina/sinb=n, where a is the angle of incidence, b is the angle of refraction and n is the refractive index.). But “in nature there is no law of refraction, only different cases of refraction” (p582). The so called law of refraction is only “a concise compendious rule, devised by us for the mental reconstruction of fact”; in fact it is only a partial reconstruction that involves idealisations etc. The appeal to unobservable entities, eg to atomic structures, can be accepted, if at all, only as a means to achieve an economical classification/systematisation. That is, Mach said, as a “mathematical model for facilitating mental reproduction of the facts” (p589). Even so, they should be admitted only as “provisional helps” and we should try to attain “more satisfactory substitute(s)”. Mach was, in fact, one of the proponents of the view that theoretical discourse in science should be systematically eliminated in favour of talk about sense experiences (what Mach called “elements”.) (“All physical knowledge can only mentally represent and anticipate compounds of those elements we call sensations. It is concerned with the connections of these elements” (p611).)

Why was the recourse to unobservables so unappealing to Mach? Last week we discussed concept empiricism and the thought that all meaningful talk must be grounded in experience. This is one important motivation for Mach. (Mach thought that the Newtonian concepts of ‘absolute space’ and ‘absolute time’ are meaningless because they are not verifiable.) Mach thought that positing entities or processes that are not observable is legitimate, but only if they play an “economical” role. He, therefore, did not advocate the naive (idealist) view that something exists iff it can be directly perceived. He thought that this principle is too restrictive for science. Here is one of his examples: Take a long elastic rod and attach it to a vise. One can strike the rod and observe its vibrations. Suppose now that one cuts the rod short and strikes it again. No vibrations are now observed. Did they cease to exist? Mach thought that if one retains the conception of vibrations, that is, if one supposes that they are still there, one can anticipate and detect several effects which should be attributed to the existence of vibrations, eg, certain tactile impressions if the rod is touched (1893, p587). Admitting the existence of non visible vibrations is then still “serviceable and economical”. It helps to systematise, organise and anticipate the phenomena. Supposing that some things exist even if they are visible “makes experience intelligible to us; it supplements and supplants experience” (p588).

But when it comes to atoms, Mach thought, things are different. Unlike vibrations, atoms cannot be perceived by the senses. Besides, when we posit atoms we have to drastically modify our experience. Mach accounted for this difference by the so-called ‘principle of continuity’. According to this: “Once we have reached a theory that applies in a particular case, we proceed gradually to modify in thought the conditions of such case, as far as it is at all possible, and endeavour in so doing to adhere throughout as closely as we can to the conception originally reached" (1893, p168). This sounds obscure. But the idea is simple: when we cut the rod short, we didn't have to modify the concept of vibration in order to apply it to the case at hand, ie, to the case of a shortened rod; we just said that vibrations became invisible. Although we couldn't observe vibrations, we could still anticipate and detect their effects in experience. But the case of atoms is different. When we move from the chemical, electrical and optical phenomena to the existence of atoms that are supposed to cause them, we invest atoms "with properties that absolutely contradict the attributes hitherto observed in bodies". The concept of atom is radically different from the concepts occurring in the phenomena that it is supposed to explain. The properties of atoms are not formed in a way continuous with the properties observed in the phenomena. For Mach this meant that "the mental artifice atom" was suspect: something to be disposed of.

Machian instrumentalism suggested that commitments to unobservable entities is not necessary (at best, they are "provisional helps") and that all physical knowledge should be ultimately expressed in terms of connections between sensations (or, in a more liberal way, between observables). We saw last week the failures associated with the idea that theoretical discourse is eliminable because it can be translated into talk about observables. Besides, the whole idea that atoms are suspect posits because they are not accessible in experience was dealt a rather serious—if not fatal—blow in the work of the French physicist **Jean Perrin** (1870-1942) who in 1913 published his Les Atomes. There he summarised his experimental work and the evidence for the reality of molecules and atoms. He cited thirteen distinct ways to calculate the precise value of Avogadro's number, that is the number of molecules contained in a mole of a gas. (Avogadro's hypothesis—dated from 1814—was that the same volumes of two different gases contain the same number of particles under the same conditions of pressure and temperature.) [The story is told in Mary Jo Nye's Molecular Reality, 1972.] **Henri Poincaré**, who was initially sceptical of the atomic hypothesis, wrote in 1912: "The brilliant determinations of the number of atoms computed by Mr Perrin have completed the triumph of atomicism. What makes it all the more convincing are the multiple correspondences between results obtained by totally different processes. Not too long ago, we would have considered ourselves fortunate if the numbers thus derived had contained the same number of digits. We would not even have required that the first significant figure be the same; this first figure is now determined; and what is remarkable is that the most diverse properties of the atom have been considered. In the processes derived from the Brownian movement or in those in which the law of radiation is invoked, not the atoms have been counted directly, but the degrees of freedom. In the one in which we use the blue of the sky, the mechanical properties of the atoms no longer come into play; they are considered as causes of optical discontinuity. Finally, when radium is used, it is the emissions of projectiles that are counted. We have arrived at such a point that, if there had been any discordances, we would not have been puzzled as to how to explain them; but fortunately there have not been any. The atom of the chemist is now a reality (...)”.

10. Duhem’s ‘Middle Way’?

Before but also after these experimental results, the existence of atoms was mostly grounded in their explanatory role: admitting the atomic hypothesis explains a number of phenomena from the chemical combinations of elements, to the kinetic theory of gases, to the Brownian motion. An interesting brand of instrumentalism arises from the denial of the claim that scientific theories aim to explain the phenomena. This is the position advocated by **Pierre Duhem**. Duhem is a very profound thinker and it’s impossible to do justice to his complex thought in a few pages. Here is only a rather quick account of his thought.

Duhem thought that the issue concerning the existence of unobservable entities belongs to the realm of metaphysics and not of science. His view of science was based on his conviction that explanation is not a concern of science but rather of metaphysics. An explanation of a set of phenomena amounts to “strip[ping] reality from the appearances covering it like a veil, in order to see the bare reality itself” (1906, p7). For him the very idea of looking behind the ‘veil of appearances’ belongs to the realm of metaphysics. Science is only concerned with experience, and as such it “is not an explanation. It is a system of mathematical propositions deduced from a small number of principles, which aim to represent as simply as completely and as exactly as possible a set of experimental laws” (p19). So physics does not aim to explain the phenomena, nor to describe the reality ‘beneath’ them. It only aims to embed the phenomena in a mathematical framework. To be sure, not in any mathematical framework, but in one that provides the simplest and most comprehensive classification. However, scientific theories are neither true nor false. In fact, only assertions that concern empirical facts can be judged with respect to their truth or falsity. As he put it: “We can say of propositions which claim to assert empirical facts, and only of these, that they are true or false” (p333). Theoretical assertions, that is assertions whose truth value cannot be determined in experience, lack truth-values. He stressed: “As for the propositions introduced by a theory, they are neither true nor false; they are only convenient or inconvenient” (p334). Being systematisations on some grounds of convenience, all that can be asserted of scientific theories is that they either square with the phenomena or they do not.

Duhem’s readers may find somewhat puzzling his occasional use of the term ‘true’ to apply to theories as a whole. For instance, in The Aim and Structure of Physical Theory, he stated the following: “Thus a true theory is not a theory which gives an explanation of physical appearances in conformity with reality; it is a theory which represents in a satisfactory manner a group of experimental laws. A false theory is not an attempt at an explanation based on assumptions contrary to reality; it is a group of propositions which do not agree with the experimental laws” (p21). The best way to understand this claim is to say that Duhem really speaks of empirical adequacy. A theory is empirically adequate iff all of its observational consequences are true, (that is, iff all it says about the observable phenomena is true). (Clearly, if a theory is empirically adequate is not necessarily true, but if it’s empirically inadequate, it is false.)

Duhem’s instrumentalism was linked to his project to lay down the foundations of physics as “autonomous science”, that is, by construction, free from any commitments to explanations of phenomena and to entities going beyond appearances (cf. pp19-21). He set out to show that if scientific theories are properly analysed and reconstructed, no commitments to explanations and hypotheses about the nature of things, eg, atomism, are needed (cf. pp304-305). (**Note**. Duhem was vehemently opposed to atomism and advocated the phenomenological programme of energetics. This was a whole theoretical framework for doing science (espoused also by Mach and Ostwald). Duhem described energetics as follows: “the principles it embodies and from which it derives conclusions do not aspire at all to resolve the bodies we perceive or the motions we resort into imperceptible bodies or hidden motions. Energetics presents no revelations about the nature of matter. Energetics claims to explain nothing. Energetics simply gives general rules of which the laws observed by experimentalists are particular cases” (1913, p183). However, energetics was nothing but a promissory note, a hope that at some point scientists will put aside the “hypothetical mechanisms” and try just to classify empirical laws, by means of principles that do not involve reference to atoms etc.)

How did Duhem react to the fact that physical theories posit all kinds of unobservable entities and claim to explain the phenomena? He suggested that theories are, in fact, divided into two parts: a) a representative (or classificatory) part, which classifies a set of experimental laws; and b) an explanatory one, which “takes hold of the reality underlying the phenomena”. Duhem understood the representative part of a theory as comprising the empirical laws and the mathematical formalism, which is used to represent, systematise and correlate these laws, while he thought that the explanatory part relates to the construction of physical models and explanatory hypotheses about the nature of physical processes which purport to simulate and reveal underlying mechanisms and causes. But, he suggested, the explanatory part is parasitic on the representative. To support this view, Duhem turned to the history of science, especially the history of optical theories and of mechanics. He argued that when a theory is abandoned because it fails to cover new experimental facts and laws, its representative part is retained, partially or fully, in its successor, while the attempted explanations offered by the theory get abandoned ("constant breaking-out of explanations which arise to be quelled" (p33).) (As we shall see next week, this is the kernel of an important argument against scientific realism, the so-called pessimistic meta-induction. In effect, Duhem's point is that the history of science is the graveyard of attempted explanations of the natural phenomena. So, one cannot warrantedly be optimistic about current explanations of the phenomena. For all we know, they too have arisen only to be subsequently quelled.)

Duhem was well aware of the fact that, with very few exceptions, the alleged two parts of a scientific theory were in fact interwoven in actual theories. How then can we distinguish between the parts that count as representative and those that are just explanatory? Duhem's basic thought seemed to be that the representative part is the mathematical expressions (laws) that encode the phenomena, while the explanatory part comprises all hypotheses concerning the causes of the phenomena. For instance, Duhem thought that Newton's law of universal gravitation typically belongs to the representative part of Newton's theory, since it "condenses" the laws of all celestial phenomena. But any attempt to characterise the cause of gravitational attraction belongs to the realm of explanation (or of metaphysics, as Duhem would put it), and as such it should not be the concern of physicists (p47). But, I think, Duhem was too quick here. One can argue that the representative/explanatory distinction is suspect. On the one hand, Duhem offers no way to tell which elements of an actual theory belong to the representative part, apart probably from a post hoc one which suggests that whatever elements were retained in theory-change belonged to the representative part of the theory. On the other hand, scientific explanations proper are representative in that they too are cast in mathematical form and, normally, entail predictions that can be tested. Take, for instance, Fresnel's explanation of the phenomenon of polarisation by means of the hypothesis that light-waves are transverse (ie, that the vibrations are executed perpendicularly to the direction of propagation.) This hypothesis was used to explain the phenomenon (discovered by Fresnel and Arago) that when two light-rays are polarised perpendicularly to one another they don't interfere, but when they are polarised parallel to each other they do. Isn't Fresnel hypothesis a legitimate scientific explanation? What is true is that the evidence doesn't logically entail (one can say, it doesn't determine) the truth of a potential explanation. But this doesn't mean that the evidence can never confirm a potential explanation (such as Fresnel's hypothesis) to a high degree. (We shall come back to this issue next week.)

Having said all this, we must also note that Duhem himself offered two important arguments against a purely instrumentalist understanding of theories. The first was that instrumentalism contradicts the scientific intuition that theories are not just catalogues of information amassed through experiments. Suppose that a physicist follows Duhem's advice to understand scientific theories as mere systematisations of empirical laws. Such a physicist would, as Duhem put it, "at once recognise that all his most powerful and deepest aspirations have been disappointed by the despairing results of his analysis. [For he] cannot make up his mind to see in physical theory merely a set of practical procedures and a rack filled with tools. (...H)e cannot believe that it merely classifies information accumulated by empirical science without transforming in any way the nature of these facts or without impressing on them a character which experiment alone would not have engraved on it. If there were in physical theory only what his own criticism made him discover in it, he would stop devoting his time and efforts to a work of such a meagre importance" (p334).

This argument against a purely instrumental reconstruction of scientific theories ("racks filled with tools") does not aim to bring out the psychological discomfort of the author of an instrumental theory who would feel that the product of his painstaking reconstruction has no cognitive value. Rather, the argument suggests that it is against scientists' pre-philosophical intuitions that the aim of a theory is not to improve our understanding of the world, but rather to classify information amassed through experiments in a convenient mathematical framework. And there seems to be nothing wrong with these intuitions which would compel scientists to change them.

Duhem's second argument was that if theories are understood as mere classifications of experimental laws, then it is difficult to explain how and why the theory succeeds in predicting novel effects. In other words, if a theory were just a "rack filled with tools", it would be hard to understand how it can be "a prophet for us" (p27). Duhem was struck by the ability of some scientific theories to predict hitherto unforeseen phenomena; eg, the prediction of Fresnel's theory of diffraction (deduced by Poisson) that if the light from a light source is intercepted by an opaque disk, then a bright spot will appear at the centre of its shadow. Duhem thought that this "clairvoyance" of scientific theories would be unnatural to expect—it would be a "marvellous feat of chance"—if "the theory was a purely artificial system" that "fails to hint at any reflection of the real relations among the invisible realities" (p28). But the same "clairvoyance" would be perfectly natural, if the principles of the theory "express profound and real relations among things". Given that theories have been successful prophets for us, if we were to bet either on theories being artificial systems or on their being "natural classifications", we would find it natural to bet on the latter. And given that we understand a theory as a natural classification, we would bet that its predictions are going to be correct. For Duhem, “the highest bet of our holding a classification as a natural one is to ask it to indicate in advance things which the future alone will reveal. And when the experiment is made and confirms the predictions obtained from our theory, we feel strengthened in our conviction that the relations established by our reason among abstract notions truly correspond to relations among things” (p28).

So, Duhem’s thought was that the fact that some theories generate novel predictions cannot be accounted for on a purely instrumentalist understanding of scientific theories. For how can one expect that an artificial classification of a set of experimental laws, ie, a classification based only on considerations of convenience, will be able to reveal unforeseen phenomena in the world? For this, it would be required that the theory has somehow “latched onto” the world; that its principles describe the mechanisms or processes that generate these phenomena. If, for instance, there were no light-waves of the kind described in Fresnel’s theory, and if the behaviour of these waves were not like the one described in this theory, how could Fresnel’s theory reveal unforeseen phenomena? Wouldn’t that be a fluke? Duhem’s conclusion was that theories that generate novel predictions should be understood as natural classifications.

How are we to understand the notion of natural classification? Duhem thought that if a theory offers a natural classification, then the relations it establishes among the experimental data “correspond to real relations among things” (pp26-27). Is this a realist position? Insofar as Duhem understood “natural classification” as revealing real relations between unobservable entities, then he defended a kind of realist position. In fact, Duhem explicitly cited Henri Poincaré’s views in support of his own. **Poincaré** thought that science can only achieve knowledge of relations between otherwise unknowable entities (“the real objects which Nature will eternally hide from us”). As he put it: “Still things themselves are not what it [i.e. science] can reach as the naive dogmatists think, but only relations between things. Outside of these relations there is no knowable reality” (1902, p28). This position may be called ‘structural realist’. It has been recently explored by Worrall and we’ll discuss it in some detail next week.)

Before we leave this section we must note that Duhem presented both of the foregoing arguments as “acts of faith”, falling totally outside the “method of physical sciences” (p27 & pp334-335). But as we shall see later, a variant of his second argument eventually became a standard realist argument against instrumentalism and was defended on the basis that scientists offer arguments of this form all the time.

11. Craig’s Theorem

Based on a theorem on first-order logic proved in early 1950's, **William Craig** proposed a clever device for an in principle elimination of all theoretical terms of a theory. Craig showed that if we can partition the vocabulary of a theory T into two classes, one theoretical the other observational, then we can syntactically replace all theoretical terms, in the sense that we can construct another theory Craig(T) which has only observational terms but is functionally equivalent with T, ie, it establishes exactly the same deductive connections between observables as T. This is the closest instrumentalism can get to the aim of a complete elimination of theoretical discourse in science. The full proof of this theorem requires more knowledge of logic than you are expected to have. (But for those of you that don't mind technical results, the proof is offered in Putnam's piece 'Craig's Theorem'.) Here I shall only indicate how this result comes about.

Craig required that two conditions be satisfied: (1) The non-logical vocabulary of the original theory T is effectively partitioned into two mutually exclusive and exhaustive classes, one containing all and only theoretical terms, the other containing all and only observational ones. Let V_T (=T_1,T_2,...,T_n) be the theoretical vocabulary and V_O (=O_1,O_2,..., O_n) be the observational one. (2) The theory T is axiomatisable, and the class of proofs in T (that is the class of applications of a rule of inference with respect to the axioms of the theory and whatever has been previously inferred from them) is effectively defined. Then Craig shows us what the axioms of the new theory Craig(T) are. There will be an infinite set of axioms (no matter how simple the set of axioms of the original theory T was), but there is an effective procedure which specifies all of them. In effect, each axiom of Craig(T) will be a very long conjunction of a single observation sentence, say O_1, conjoined to itself many times (ie, O_1&O_1&O_1 &...&O_1). (**Note**. The number of times an observation sentence O appears in the relevant axiom of Craig(T) is the Gödel number of a proof of O itself in the original theory T.) The new theory Craig(T) which replaces the original theory T is 'functionally equivalent' to T, in that all observational consequences of T also follow from Craig(T). In other words, T and Craig(T) have exactly the same empirical content. Here is a very simple example. Take a theory T such that an observation sentence O_1 implies another observation sentence O_2 by virtue of a theoretical sentence T_1, ie, let O_1 □ T_1 imply O_2. Then T_1 implies the (observational) conditional $O_1 \varnothing O_2$. The Craig(T) will clearly not have T_1. But since $O_1 \varnothing O_2$ is an observation sentence, it will belong to Craig(T), and so Craig(T) will also entail the prediction O_2, since (O_1 □ ($O_1 \varnothing O_2$)) implies O_2. Hence, Craig(T) establishes all those deductive connections between observation sentences that the initial theory T establishes. The general moral is that, for any V_O-sentence O_0, if T implies O_0 then Craig(T) implies O_0. It is in this sense that the Craig(T) is functionally equivalent with T (cf. Hempel, 1958, pp75-76 & 1963, p699).

Instrumentalists have always argued that theoretical commitments in science are dispensable. Craig’s theorem offers a boost to instrumentalism, by proving that theoretical terms can be eliminated en bloc, without loss in the deductive connections between observables established by the theory. Carl Hempel (1958, pp49-50) presented the significance of Craig’s theorem in the form of dilemma, the theoretician’s dilemma: If the theoretical terms and the general principles of a theory don’t serve their purpose of a deductive systematisation of the empirical consequences of a theory, then they are surely unnecessary. But, given Craig’s theorem, if they serve their purpose, “they can be dispensed with since any chain of laws and interpretative statements establishing such a connection should then be replaceable by a law which directly links observational antecedents to observational consequents”. But the theoretical terms and principles of a theory either serve their purpose or they don’t. Hence, the theoretical terms and principles of any theory are unnecessary.

How can one react to Craig’s theorem? There are three general ways to dispute the significance of Craig’s theorem. First, as we saw in the last lecture, there is no principled way to draw a line between theoretical and observational terms. Since Craig’s theorem requires a separation of the language of a theory in two vocabularies, its worth is as good as the claim that we can divide the language of science into a theoretical and an observational vocabulary. Second, although from a logical point of view theoretical terms are dispensable, if they are dispensed with, then scientific theories will lose several of their salient features, eg, simplicity and predictive fertility. In particular: (a) If the Craig(T) replaces a theory T, then the comparative simplicity of a theoretical system will be lost. Craig(T) will always have an infinite number of axioms, no matter how few, simple and elegant the axioms of the original theory were. As Hempel put it: “this price is too high for the scientist, no matter how welcome the possibility of the replacement may be to the epistemologists”. (b) If one replaces T with Craig(T), then scientific theories may lose in predictive fertility and heuristic power. Suppose we have two theories T_1 and T_2 which have the same observational and theoretical vocabularies and which are consistent (individually and jointly). Suppose also that we conjoin T_1 and T_2 to form the theory T_1&T_2. Generally, T_1&T_2 will entail some extra observational consequences that neither T_1 alone nor T_2 alone would entail. If we had replaced T_1 by Craig(T_1) and T_2 by Craig(T_2), and had then formed the conjunction Craig(T_1)&Craig(T_2), we might have failed to generate the extra observational consequences. (The consequences of combining Craig(T_1) and Craig(T_2) are a proper subset of the observational consequences of combining T_1 and T_2.) The point of this is that the original theory T has a potential over time over Craig(T): the theoretical terms of T may help the generation of new observational predictions which cannot be generated with the help of Craig(T) alone.

The third response to Craig's theorem is to explicitly deny the claim that theories are just vehicles for deductive transitions among observable sentences. As Hempel (1958, p78) argued, theories should also serve "inductive systematisations" of phenomena (that is, they should establish inductive connections between observations. Some of the inductive links that the theory establishes would be unattainable without the use of theories and theoretical terms. Therefore, even if a theory T and its Craig(T) are "functionally equivalent" vis-à-vis the deductive connections between observables, this 'equivalence' breaks down when inductive relations between observables are taken into account. Here is Hempel's own example: Take a simple theory whose only theoretical terms are 'white phosphorus' ('P') and 'ignition temperature of 30^{o} C' ('I'). The theory has two general principles: 'White phosphorus ignites at a temperature of 30^{o} C', and 'When the ignition temperature is reached, phosphorus bursts into flames'. Let's express these two principles as follows: (1) $\forall x\ (Px \varnothing Ix)$; (2) $\forall x\ (Ix \varnothing Fx)$. Suppose now that we know of certain necessary conditions for the presence of white phosphorous, eg, that white phosphorous has a garlic-like odour ('G'); it is soluble is turnpentine ('T'), in vegetable oils ('V') and in ether ('E'); it produces skin burns ('S'). Let's express them in symbolic form as follows: (3) $\forall x\ (Px \varnothing Gx)$; (4) $\forall x\ (Px \varnothing Tx)$; (5) $\forall x\ (Px \varnothing Vx)$; (6) $\forall x\ (Px \varnothing Ex)$; (7) $\forall x\ (Px \varnothing Sx)$. Let all these seven sentences represent the total content of the theory T. Clearly, principles (1) and (2) above do not have any observational consequences and hence they cannot be used for the relevant deductive systematisation. However, these principles can be used to establish inductive connections between observables. Suppose that a certain object **b** has been found to have a garlic-like odour, to be soluble in turnpentine, in vegetable oils and in ether, and to produce skin burns. Then, one can use sentences (3) to (7) to inductively conclude that **b** is white phosphorous. One can then use principles (1) and (2) above to infer that **b** will burst into flames if the temperature reaches 30^{o} C. That is, one can derive a certain observational prediction that could not be derived without the inductive transition from certain observational sentences 'Gb', 'Tb', Vb', Eb', 'Sb', via sentences (3) to (7), to the theoretical claim that the object under investigation is white phosphorous, ie, 'Pb'. The same inductive transition could not have been made if one had replaced the original theory by its Craig-transform.

More generally, a theory T may be said to help establish inductive connections between observables, if there are observation sentences O_1 and O_2 such that Prob(O_1/O_2&T) > >Prob(O_1/O_2), or, more interestingly, if Prob(O_1/O_2&T) > >Prob(O_1/O_2&Craig(T)). Let me illustrate how this can happen by means of an example (due to Putnam). Imagine that you are involved in a nuclear fusion experiment and you consider the prediction H: when two subcritical masses of U_{235} are slammed together to from a supercritical mass, there will be a nuclear explosion. A variant of H can be stated in a purely observational language, that is without the term 'Uranium 235', as O_1: when two particular rocks are slammed together, there will be an explosion. Consider now the available observational evidence, namely that O_2: up to now, when two rocks of a particular sort were put together nothing happened. It follows that Prob(O_1/O_2) is very low. Nevertheless, you are confident of H. Why is that so? Because, one may say, Prob(H/O_2 & Atomic Theory)>> Prob(H/O_2). (Remember, H is the theoretical variant of O_1.) More interestingly, one may say, Prob(H/O_2 & Atomic Theory)>> Prob(H/O_2 &Craig(Atomic Theory)). The claim here is that the atomic theory makes it likely that the two Uranium_{235} rocks will explode if critical mass is attained quickly enough. But if we had a theory that involves no relevant theoretical statements would we have any basis to expect that the two rocks will explode? If we had only considered the Craig(Atomic Theory), would we have the same confidence in H? (Notice here that we don't deny that the atomic theory can be replaced by its Craig-transform. The point of this example is that if we eliminate theories proper, we may miss out in theory-guided hitherto unforeseen inductive connections between observables).

Based on the last reaction to Craig's theorem, Hempel himself dismissed the paradox of theorising by saying that it "starts with a false premiss" (1958, p87). In other words, theories are not just a deductive systematisation of data, after all, and there is no compelling reason to be thus interpreted, without also irredeemably losing some of their fruitfulness.

Putnam, once again, pushed this line to its extremes by challenging the then dominant view of the aim of scientific theories, viz., that ultimately the purpose of theories is 'prediction and control'. He argued that if scientists employ terms like 'electron', 'virus', 'space-time curvature' and so on—and advance relevant theories—it is because they wish to speak about electrons, viruses, the curvature of space-time and so on; that is they want to find out about the unobservable world. But then how can we eliminate the use of theoretical terms? After all, it is these terms that provide us with the necessary linguistic tools for talking about things we want to talk about.

12. Motivations for Realism

So far we've examined several forms of instrumentalism and have described their limitations. But aren't there positive arguments for a realist understanding of scientific theories? Let me first clarify what exactly a realist understanding of theories involves. It involves two theses: one semantic/ontological, and another epistemic.

1) Scientific theories should be taken at face-value. They are truth-valued descriptions of their intended domain, both observable and unobservable. Theoretical assertions are not reducible to claims about the behaviour of observables, nor are they merely instrumental devices for establishing connections between observables. The theoretical terms featuring in them have putative factual reference.

2) Mature and predictively successful scientific theories are approximately true of the world; the entities posited by them—with the structure, properties and causal behaviour attributed to them—inhabit the world.

(Different authors may have different understandings of what scientific realism should involve. The above is the strongest realist position. It asserts that scientific theories make ineliminable assertions about the unobservable world and that they can offer us knowledge of this world. The first thesis is now more or less generally accepted. Current debates revolve mostly around the second, epistemic thesis. In the next lecture, we shall see to what extent this thesis is defensible.)

The major argument for realism has been that it offers a much more plausible understanding, or explanation, of the fact that the observable world is as scientific theories describe it to be. This is a line that was developed by many realists. **Jack Smart**, for instance, described instrumentalism [phenomenalism] as the view that “statements about electrons, etc., are only of instrumental value: they simply enable us to predict phenomena on the level of galvanometers and cloud chambers” (1963, p39). Then he stressed that “if the phenomenalist about theoretical entities is correct we must believe in cosmic coincidence”. Instrumentalists who claim that theoretical assertions have only instrumental value, or explicitly deny the existence of unobservable entities posited by scientific theories, are committed to a gigantic cosmic coincidence: the phenomena just happen to be the way scientific theories say they are; they just happen to be related to one another the way scientific theories say they are. But realism leaves no space for such happenstance: it’s because theories are true that the phenomena are the way scientific theories describe them. As Smart put it: “Is it not odd that the phenomena of the world should be such as to make a purely instrumental theory true? On the other hand, if we interpret a theory in the realist way, then we have no need for such a cosmic coincidence: it is not surprising that galvanometers and cloud chambers behave in the sort of way they do, for if there are really electrons, etc., this is just what we should expect (p39).

Grover Maxwell was probably the first to focus on the claim that the success of scientific theories is a fact that calls for an explanation. He offered an argument for realism by saying that: “The only reasonable explanation for the success of theories of which I am aware is that well-confirmed theories are conjunctions of well-confirmed, genuine statements and that the entities to which they refer, in all probability exist” (1962, 18). On the contrary, claiming that scientific theories are ‘black boxes’ which are fed with observational premisses and yield observational conclusions would offer no explanation of the fact that these ‘black boxes’ work as well as they do. As he later on pointed out, the difference between realism and all other philosophical accounts of science is that “as our theoretical knowledge increases in scope and power, the competitors of realism become more and more convoluted and ad hoc and explain less than realism. For one thing, they do not explain why the theories which they maintain as mere, cognitively meaningless instruments are so successful, how it is that they can make such powerful, successful predictions. Realism explains this very simply by pointing out that the predictions are consequences of the true (or close true) propositions that comprise the theories” (1970, 12).

We saw already that Duhem too thought that the novel predictive success of science is in conflict with a purely instrumentalist understanding of theories. But he thought that such a ‘plausibility argument’ falls outside the scope of scientific method. Maxwell, however, suggested that the defence of realism on the grounds that it best explains the success of science has the same form as the arguments that scientists use to support their own theories. Scientists normally suggest that the explanatory successes of a theory confirm the theory. Similarly, philosophers can argue that the predictive and explanatory success of science as a whole confirms realism. Realism is not defended as an a priori truth. Rather, it is defended as a contingent theory that gets supported from the success of science. The claim is that the success of science makes realism much more probable than instrumentalism. In point of fact, Maxwell turned his argument into a straightforward Bayesian. Suppose that both realism (R) and instrumentalism (I) entail the success of science (S). Then, the likelihoods of realism and instrumentalism are both equal to unity, ie, $Prob(S/R)=Prob(S/I)=1$. Then the posterior probability of realism is $Prob(R/S)=Prob(R)/Prob(S)$ and the posterior of instrumentalism is $Prob(I/S)=Prob(I)/Prob(S)$, where $Prob(R)$ is the prior probability of realism, $Prob(I)$ is the prior of instrumentalism and $Prob(S)$ is the probability of the ‘evidence’, ie, of the success of science. But, Maxwell argued, given that the prior probability of realism is much greater than the prior of instrumentalism (ie, $Prob(R)>>Prob(I)$), the confirmation of realism is much greater than that of instrumentalism. Maxwell argued that in science “prior probability is always one of the crucial factors in selecting among competing hypotheses, all of which explain current evidence”. He just applied this in the case of competing hypotheses concerning scientific theories and defended realism as “more highly confirmed than any other theory that ‘explains the facts’” (1970, p17). (But how can we defend the claim that the prior probability of realism is much higher than that of instrumentalism? Discuss in class.)

Putnam turned this line into the most famous slogan for scientific realism, the so-called ‘no miracle’ argument :

“The positive argument for realism is that it is the only philosophy that doesn’t make the success of science a miracle. That terms in mature scientific theories typically refer (this formulation is due to Richard Boyd), that the theories accepted in a mature science are typically approximately true, that the same terms can refer to the same even when it occurs in different theories—these statements are viewed not as necessary truths but as part of the only scientific explanation of the success of science, and hence as part of any adequate description of science and its relations to its objects” (1975, p73).

This seems a powerful argument. After all, one may think, is it not clear that realism offers the best explanation of the success of science? Does instrumentalism offer any explanation at all? If instrumentalism is correct, isn’t the novel predictive success of theories totally unexpected? And is it not reasonable to adopt the best explanation of the evidence. Isn’t this what scientists do all the time? But this argument has been challenged on several grounds. First, this argument begs the question in favour of realism. For it is itself an inference to the best explanation and non-realists deny that inference to the best explanation is a reasonable inferential method. Second, the realist explanation of the success of science is destroyed by the history of science. For this history is full of theories that were once empirically successful and yet typically false. Third, even if we grant that the argument is not question-begging, there are better non-realist potential explanations of the success of science. We’ll discuss all these objections next week.

There is however a more immediate challenge to the realist argument. It is this. Let us grant that this argument is effective against traditional instrumentalism. That is, let us grant that it is more probable that an unobservable world exists than that nothing but observable phenomena exist. One could still take a sceptical attitude towards the theoretical descriptions of the world offered by scientific theories. That is, one could argue that we do not have, or indeed could not have, good reasons to believe that the particular theoretical descriptions of the world are true. This is the line that **Bas van Fraassen** takes. He is not an instrumentalist of the traditional sort. In fact, he countenances the realist position (1) above. But he is an agnostic instrumentalist, or a sceptic. His main point is that given that all kinds of different theoretical descriptions of the world fit the observable phenomena, we have no reason to believe in one rather than any other. The probability of the observable phenomena being what they are, given that current theories T are true, is not greater than the probability of the observable phenomena being what they are, given that current theories T are false; that is, given that the unobservable world is different from the way our current theories describe it. We are then better off if we remain agnostic about truth and opt for empirical adequacy. Van Fraassen's agnosticism is based largely on the so-called argument from the underdetermination of theories by evidence. We shall discuss this in some detail next week. For the time being, let us just point out that even if traditional instrumentalism is discredited, some agnostic versions of it seem to survive.

Study Questions

1. Outline Duhem's position concerning scientific theories and his arguments against the view that theories are just "racks filled with tools".

2. What is Craig's Theorem? a) How does it motivate an instrumentalist account of scientific theories? b) How can one respond to it?

3. Carefully state Hempel's theoretician's dilemma. How does it relate to Craig's theorem? How did Hempel react to it? (Work on Hempel's piece "The Theoretician's Dilemma", pp67-71 & p75-87.)

4. What does a realist understanding of scientific theories involve?

5. Present and critically discuss the 'no miracle' argument for realism. How did Maxwell try to defend realism?

Further Reading

Duhem, P. (1906) The Aim and Structure of Physical Theory, second edition 1914, translated by P. Wiener 1954, Princeton University Press.

Duhem, P. (1908) To Save the Phenomena, E. Doland & C. Mascher (trans.), 1969, The University of Chicago Press.

Duhem, P. (1913) 'Examen Logique de la Théorie Physique', translated in English as 'Logical Examination of Physical Theory', P. Barker & R. Ariew (trans.), Synthese, (1990), **83**, pp.183-188.

Hempel, C. (1958) 'The Theoretician's Dilemma: A study in the Logic of Theory Construction', Minnesota Studies in the Philosophy of Science, **2**, University of

Minnesota Press.

Mach, E. (1893) The Science of Mechanics, T J McCormack (trans.) Sixth Edition, Open Court Publishing Company.

Maxwell, G. (1962) ‘The Ontological Status of Theoretical Entities’, Minnesota Studies in the Philosophy of Science, **3**, University of Minnesota Press.

Maxwell, G. (1970) ‘Theories, Perception and Structural Realism’, in R Colodny (ed.) The Nature and Function of Scientific Theories, Pittsburgh: University of Pittsburgh Press.

Putnam, H. (1965) ‘Craig’s Theorem’ in Putnam, Mathematics, Matter and Method.

Putnam, H. (1975) Mathematics, Matter and Method, Philosophical Papers Vol.1, Cambridge University Press.

Smart, J. J. C. (1963) Philosophy and Scientific Realism, London: RKP.

IV. Scientific Realism and its Critics

13. The Explanationist Defence of Realism

A main motivation for scientific realism is **the ‘no miracle’ argument**: realism is the only theory of science that doesn’t make its success a miracle. The claim is that empirically successful scientific theories should be accepted to be approximately true, for otherwise their success remains unaccounted for. This line of thought has been developed by **Richard Boyd** into a whole philosophical programme for the defence of a realist epistemology of science, what Putnam once called “the Best Explanation epistemology”—also known as the explanationist defence of realism (EDR). Boyd defends realism on the basis that it offers the best explanation of the instrumental success of science. Boyd’s argument for realism is an instance of the so-called **inference to the best explanation** (or, **abduction**) and its aim is to show that realism can rationally be defended because it offers the most coherent, comprehensive and potentially explanatory understanding of science. (**Note**. Inference to the Best Explanation is the mode of inference in which we infer to the truth of one of a set of inconsistent potential explanatory hypothesis on the basis that the chosen hypothesis provides the best explanation of the evidence.) Let’s see Boyd’s argument in some more detail.

Boyd has set out to show that the best explanation of the instrumental and predictive success of mature scientific theories is that these theories are approximate true, <u>at least in the respects relevant to the instrumental success</u>. Boyd’s novelty is that he made the generally accepted theory-ladenness of scientific methodology central to the defence of realism. Here is a reconstruction of his argument.

> There is a general consensus over the claim that the methods by which scientists derive and test theoretical predictions are theory-laden. Scientists use accepted background theories in order to form their expectations, to choose the relevant methods for theory-testing, to devise experimental set-ups, to calibrate instruments, to assess the experimental evidence, to choose a theory etc. All aspects of scientific methodology are deeply theory-informed and theory-laden. In essence, scientific methodology is almost linearly dependent on accepted background theories: it is these theories that make scientists adopt, advance or modify their methods of interaction with the world and the procedures they use in order to make measurements and test theories.
>
> These theory-laden methods lead to correct predictions and experimental successes. (They are instrumentally reliable.)

> How are we to explain this?
>
> The best explanation of the instrumental reliability of scientific methodology is this: the statements of the theory which assert the specific causal connections or mechanisms in virtue of which methods yield successful predictions are approximately true.

Let me illustrate the conclusion by means of a general example. Suppose, for instance, that a theory T says that method M is reliable for the generation of effect X in virtue of the fact that M employs causal processes $C_1,...,C_n$ which, according to T, bring about X. Suppose, also, that one follows M and X obtains. Boyd's conclusion is that the best explanation of getting X is that the theory T—which asserted the causal connections between $C_1,...,C_n$ and X—is approximately true. (Think of the theory-led methods that produce effective cures to diseases. Or, for that matter, think of Fresnel's wave-theory of light. As we already saw, it suggests that if light-waves are intercepted by an opaque circular disk, then the light-waves get diffracted at the edge of the disk and they produce a bright spot at the centre of its shadow. What is the best explanation of the ability of Fresnel's theory to yield this successful prediction? Isn't it that Fresnel's theory is (approximately) true? As Boyd once put it: "The only scientifically plausible explanation of the reliability of a scientific methodology which is so theory-dependent is a thorouhgoingly realistic explanation: Scientific methodology, dictated by currently accepted theories, is reliable at producing further knowledge precisely <u>because, and to the extent that, currently accepted theories are relevantly approximately true</u>". For Boyd theory and method in fact blend in one dialectic relationship: the approximate truth of accepted theories explains why our theory-dependent methodology is reliable and that reliable methodology, in turn, helps to produce new theoretical knowledge by showing how to improve our current theories.

Boyd's argument aims to go far beyond the original 'no miracle' argument. As we saw last week, the original 'no miracle' argument was not effective against **agnostic** versions of instrumentalism. For one could go along with the 'no miracle' argument as far the <u>existence</u> of an unobservable world is concerned and yet remain agnostic as to the <u>correctness</u> of any particular theoretical description of it. In other words, one could acknowledge that theories make ineliminable reference to an unobservable world (ie, their theoretical commitments are not dispensable) and yet simply remain <u>agnostic</u>, ie, suspend judgement, on whether current theories are true or false of this world. (This is, as we shall explain in more detail later, the position advocated by **van Fraassen**.) Boyd's important contribution to the debates over scientific realism has been that he's tried to block this agnostic attitude. He's tried to defend the view that current successful theories can warrantedly be accepted as approximately true. So, not only is it the case that theoretical commitments in science are ineliminable but also the theoretical commitments issued by current theories are (approximately) true. Hence, Boyd has tried to defend the full-blown realist position described in the last section of the previous lecture.

Let's point out some features of Boyd's argument. It rests on the idea that it is rational to believe the best explanation of the evidence. In fact, it is meant to be an argument that defends inference to the best explanation as a legitimate and reliable mode of inference. As we saw, the argument suggests that the best explanation of the instrumental success of scientific methodology is that background theories are (approximately) true. These background scientific theories, however, have been typically the product of inference to the best explanation. Hence, hypotheses that are generated by means of IBE tend to be true. In essence, Boyd's argument suggests that in the light of the fact that scientific methodology is theory-laden, the best explanation of the instrumental success of this methodology is that reasoning to the best explanation is legitimate and reliable.

There is an immediate criticism to the above argument, viz., that it is circular and question-begging. Let's carefully analyse the critics' complaint. It should be clear that Boyd's argument is itself an instance of an inference to the best explanation (IBE). It defends realism by arguing that it is the best explanation of the instrumental reliability of science. But then it is not difficult to see that this argument is circular. For, in essence, it uses inference to the best explanation in order to defend inference to the best explanation. The argument uses IBE in order to defend realism; but realism itself involves the thesis that IBE is a legitimate inferential method, ie, that it is rational to believe a hypothesis on the grounds that it best explains the evidence. But, the critics go on, the issue at stake between realists and non-realists is precisely whether we should believe a hypothesis on the grounds that it best explains the evidence. Realists affirm this, while non-realists deny it. But then, Boyd's argument begs the question against them: it presupposes what they deny. As **Arthur Fine** put it, it employs "the very type of argument whose cogency is the question under discussion" (Fine, 1991, p82), and as **Larry Laudan** put it, the 'no miracle' argument is "the realists' ultimate petitio principii" (1984, p134).

How can realists reply to this objection? This is a rather delicate issue that will take us into deep waters. (The interested reader should look at Boyd's own papers—especially 'The Current Status of the Scientific Realism', section on 'Issues of Philosophical Method' and at **David Papineau**'s Philosophical Naturalism, chapter 5.) I will only very briefly indicate some possible answers at the realist's disposal. First, realists admit that the argument is circular, but they qualify the notion of circularity so that not all circular arguments are vicious and question-begging. The claim then is that Boyd's argument is circular but not viciously so. Problems with this suggestion: are there any non vicious circles?; the suggestion proves 'too much'. Second, realists admit that the argument is circular, but note that, more or less, all philosophical positions reach their conclusions via arguments that their rival will think unjustified and question-begging. The issue is not to get involved in a sterile debate about how to avoid circularity, but rather to assess the rival positions with respect to their overall adequacy, comprehensiveness etc. Problems with this suggestion: on what basis are we to assess overall adequacy?; stand-off between the rivals. Third, realists argue that the issue of circularity is an endemic problem that characterises all attempts to justify an ampliative rule of inference. (But not only those. The issue of circularity also arises when we try to justify deduction.) However, they say, those non-realists that are not fully sceptical do recognise that although inductive learning cannot be defended in a way that solves Hume's problem, it offers means to go beyond observed associations and form warranted beliefs about unobserved patterns. In this respect, inference to the best explanation is no better or no worse than induction and inductive learning from experience. As in the case of induction, it cannot be defended by a non-circular argument, but it offers means to go beyond observed associations and form warranted beliefs about unobservable patterns. Some non-realists complain that there is a special problem with the knowledge of unobservables. Yet, realists reply, they have not shown us that the unobservable should be identified with the epistemically inaccessible. Problems with this suggestion: no positive reason to trust abduction; stand-off with the sceptic about unobservables. Realists can go for a combination of the foregoing answers. **BUT** it's worth noting that there is a stand-off between realists and their critics: realists cannot offer non-realists positive reasons to trust IBE, but non-realists have not succeeded in undermining IBE either. (For more details on this one can see my paper 'On van Fraassen's critique of abductive reasoning'.)

14. Are There Better Explanations of the Success of Science?

Be that as it may, realists need to show that their potential explanation of the success of science is better than its rivals. But is this the case? Let's examine a few rivals. Arthur Fine has aimed to show that versions of instrumentalism can offer a better explanation of the success of science. He suggests that some notion of pragmatic (or instrumental) reliability of scientific theories best explains the success of science, where "instrumental reliability" is a feature of scientific theories in virtue of which they are "useful in getting things to work for the practical and theoretical purposes for which we might put them to use" (1991, p86). He contrasts two forms of (simplified) abductive reasoning from the success of science (1986a, pp153-154; 1991, pp82-83):

(A)	(B)
Science is empirically successful	Science is empirically successful
∴ (Probably) Theories are instrumentally reliable	∴ (Probably) Theories are approximately true

Fine suggests that pattern (A) is always preferable to (B) on the grounds that if the explanandum is the instrumental success of scientific methodology, we do not need to inflate the explanans with “features beyond what is useful for explaining the output” (1991, p83). So, he points out, “the instrumentalist, feeling rather on home ground, may suggest that to explain the instrumental success we need only suppose that our hypotheses and theories are instrumentally reliable” (1991, pp82-83).

But why does an appeal to the (approximate) truth of background scientific theories go beyond the features that are useful for explaining instrumental success? Fine’s suggestion is the following. When a realist attempts to explain the success of a particular theory she appeals to the truth of a theoretical story as the best explanation of the theory’s success in performing several empirical tasks. But if this explanation is any good at all, she must “allow some intermediate connection between the truth of the theory and success in its practice. The intermediary here is precisely the pragmatist’s reliability” (1986a, p154). Then, he suggests, the job that truth allegedly does in the explanation of the success of a theory is in fact done by this intermediate pragmatic reliability. Truth seems explanatorily redundant. Moreover, if pragmatic reliability is substituted for truth in the realist account of success, one gets the alternative account in terms of instrumental reliability (ibid.). Fine concludes, “since no further work is done by ascending from that intermediary to the realist’s ‘truth’, the instrumental explanation has to be counted as better than the realist one. In this way the realist argument leads to instrumentalism’ (ibid.). Moreover, on the basis of this argument, Fine suggests a meta-theorem: “If the phenomena to be explained are not realist-laden, then to every good realist explanation there corresponds a better instrumentalist one” (ibid.)

Let’s take a close look at Fine’s argument. The realists can argue that it is not at all obvious that there is anything like a pragmatic notion of reliability that realists have to take into account in their explanation of the success of science. Nor is it obvious that such a pragmatic notion of reliability intervenes between theory and experimental success.

Clearly, the realist can say, between successful empirical results and theories there are methods, auxiliary assumptions, approximations, idealisations, models, etc. Let us suppose that all these are what Fine calls pragmatic intermediary. Let us also suppose that these things alone can account for the empirical success of a theory. Would this fact make claims concerning the truth of the theory explanatorily superfluous? Surely not. For one also wants to know why some particular models work whereas others don't, or why a model works better than others, or why the methods followed generate successful predictions, or why some idealisations are better than others and the like. Then, realists can explain the successful constraints theories place on model-construction as well as the features of scientific methods in virtue of which they produce successful results by means of the claim that background theories are approximately true. But then approximate truth is not explanatorily redundant.

More importantly, suppose that we grant that there is some other pragmatic or instrumental notion of reliability to be interpolated between claims of approximate truth and claims of empirical success. Can this offer any explanation of the success of science?   In essence, instrumental reliability is nothing over and above the ability of a theory to successfully perform practical tasks. If we then explain the theory's instrumental success by saying that background theories are instrumentally reliable, it is as though we're saying the same thing in different words. That is, whether one says that theories are successful or one says that they are instrumentally reliable, one says the same thing. One doesn't provide an explanation of the theories' success; one merely rephrases success as instrumental reliability. The situation here is totally analogous with 'explaining' the fact that hammers can be successfully used to knock nails in the wall by saying that hammers are instrumentally reliable for nail-knocking! Let's recall that what is at stake here is whether an instrumentalist explanation of the success of science is better than the realist. It turns out that it doesn't seem to be an explanation at all.

Fine has recently suggested a way to make claims of instrumental reliability potentially explanatory. He outlined a dispositional understanding of the instrumental reliability of science. On this view, instrumental reliability involves a disposition to produce correct empirical results. Fine claimed that an explanation of the success of science in terms of this dispositional account of instrumental reliability is "an explanation of outcomes by reference to inputs that have the capacity (or "power") to produce such [i.e., instrumentally reliable] outcomes" (1991, p83). Clearly, this understanding of instrumental reliability is potentially explanatory, for it accounts for empirical success by an appeal to a capacity, or disposition, that theories have in virtue of which they are empirically successful. This account, however, seems problematic, too. Not because there are no dispositions or powers in nature, but rather because one would expect an explanation of why and how theories have such a disposition to be instrumentally reliable; in particular an explanation that avoids the troubles of Molière's account of why opium sends somebody to sleep in terms of the 'dormitive power' of opium. It seems natural to suggest that an explanation of this disposition in terms of the categorical property of being approximately true would ground the power of scientific theories to be instrumentally reliable. However, this would make the realist explanation better than the instrumentalist.

What about **van Fraassen**'s Darwinian explanation of the success of science? Isn't it better than the realist ? Van Fraassen's story is this: "The success of science is not a miracle. It is not even surprising to the scientific (Darwinist) mind. For any scientific theory is born into a life of fierce competition, a jungle red in tooth and claw. Only the successful theories survive—the ones which in fact latched on to actual regularities in nature" (1980, p40). That is, van Fraassen says, there is no surprise in the fact that current theories are empirically successful. For the Darwinian principle of the survival of the fittest has operated. Current theories have survived because they were the fittest among their competitors—fittest in the sense of latching onto universal regularities. Clearly, this is an elegant and simple explanation of the fact that current theories are successful. But does it undermine the realist explanation? If we unpack van Fraassen's story we find that it is a phenotypic one: it provides an implicit selection mechanism according to which entities with the same phenotype, ie, empirical success, have been selected. But of course a phenotypic explanation does not exclude a genotypic one: an explanation in terms of some underlying features that successful theories share in common, features that made them successful in the first place. The realist explanation in terms of truth provides this sort of genotypic explanation: every theory that possesses a specific phenotype, ie, it is empirically successful, also possesses a specific genotype, ie, approximate truth, which yields this phenotype. In order to see the point more clearly, compare van Fraassen's story with this: A group of people have all red hair. That's not a surprise. It is explained by the fact that they are all members of the club of red-haired persons. (The Club is, in sense, a selection mechanism, that allows only persons with red hair.) But this observation does not explain why George (or, for that matter, anyone of them taken individually) has red hair. A different, most likely genetic, story should be told about George's colour of hair.

Notice here that the realist explanation is compatible with van Fraassen's Darwinian one. Yet, the realist is arguably preferable, because it is deeper. It does not stay on the surface—that is, it does not just posit a selection mechanism which lets through only empirically successful theories. It rather tells a story about the deeper common traits in virtue of which the selected theories are empirically successful. As **Peter Lipton** (1991, p170ff.) has suggested, there is another reason for preferring the genotypic explanation to the Darwinian one. It is this. All that the phenotypic explanation warrants is that theories that have survived through the selection mechanism have not been refuted yet. There is no warrant that they will be successful in the future. Any such warrant must be external to the phenotypic story. For instance, this warrant can come from a combination of the phenotypic explanation with the principle of induction. On the other hand, the genotypic explanation has this warrant in its sleeve: if a theory is empirically successful because it is true, then it will keep on being empirically successful.

To sum up, although the debate still goes on, there seem to be no better explanations of the success of science than the realist one.

15. The Pessimistic Meta-Induction

However, the realist 'no miracle' argument has a very serious rival that seems to destroy its credibility: the history of science. Larry Laudan has argued that the history of science destroys the realist explanation of the success of science. For it is full of cases of theories that were once empirically successful and yet they turned out to be false. Laudan's argument against scientific realism is simple but powerful (cf. 1981, pp32-33; 1984, pp91-92; 1984a, pp121; 1984b, p157). It can be stated in the following form:

> The history of science is full of theories which had been empirically successful for long periods of time and yet have proved to be false about the deep-structure claims they had made about the world. It is similarly full of theoretical terms featuring in successful theories which don't refer. Therefore, by a simple (meta-)induction on scientific theories, our current successful theories are likely to be false (or, at any rate, more likely to be false than true), and many or most of the theoretical terms featuring in them will turn out to be non-referential.

Mary Hesse put the same thought in the form of the 'Principle of No Privilege', which, she said, follows from an "induction from the history of science". According to this, "our own scientific theories are held to be as much subject to radical conceptual change as past theories are seen to be" (1976, p264).

Laudan substantiated his argument by means of what he called "the historical gambit": the following list—which "could be extended ad nauseam"—gives theories which were once empirically successful and fruitful, yet neither referential nor true: they were just false.

Laudan's list of successful-yet-false theories

- the crystalline spheres of ancient and medieval astronomy;
- the humoral theory of medicine;
- the effluvial theory of static electricity;
- catastrophist geology, with its commitment to a universal (Noachian) deluge;
- the phlogiston theory of chemistry;
- the caloric theory of heat;
- the vibratory theory of heat;
- the vital force theory of physiology;
- the theory of circular inertia;
- theories of spontaneous generation;
- the contact-action gravitational ether of Fatio and LeSage;
- the optical ether;
- the electromagnetic ether.

If Laudan is right, then the realist's explanation of the success of science flies in the face of the history of science. The history of science cannot possibly warrant the realist belief that current successful theories are approximately true, at least insofar as the warrant for this belief is the 'no miracle' argument.

How can realists respond to this argument? Clearly, if realism is to be defended the general strategy of the realist response should be either to show that Laudan has overstated his case against realism or to try to reconcile the historical record with some form or other of the realist claim that successful theories are typically approximately true. (**Note**. Realists should immediately concede that Laudan's argument shows that the truth and nothing but the truth cannot be had in science. Most theories are, strictly speaking, false. Yet, they argue that false theories can nonetheless be approximately true. The problem with this suggestion is that there is no formal and objection-free way to explicate the notion of approximate truth. So non-realists complain that the semantics of approximate truth is unclear. Realist reply that there is an intuitively clear understanding of approximate truth and that's all they need in their defence of realism. A theory is approximately true is it makes roughly the right claims about posits roughly like the entities that populate the world. So, for instance, if there are entities in the world pretty much like what the theory describes as electrons, then this theory is said to be approximately true of electrons. **Ernan McMullin** put the realist position as follows: "calling a theory 'approximately true', then would be a way of saying that entities of the general kind postulated by the theory exist. It is 'approximate' because the theory is not definitive as an explanation; more has to be said.

But it already has a bearing on the truth because we can say that it has allowed us discover that entities of a certain sort exist, entities that we could not (for the moment at least) have known without the aid of the theory" (1987, pp59-60) Here again, we are on a subtle matter and I'll let you judge whether realists are entitled to appeal to some intuitive notion of approximate truth.)

Let us analyse the structure of the 'pessimistic induction' and try to see the moves that are available to realists. Laudan's argument is meant to be a kind of reductio. The target is the realist thesis that: (A) Current successful theories are approximately true. Note, however, that Laudan doesn't directly deny that current successful theories may happen to be approximately true. The aim of the argument is to discredit the realist potential warrant for such a claim, viz., the 'no miracle' argument. In order to achieve this, the argument proceeds by comparing a number of past theories to current ones and claims: (B) If current successful theories are accepted to be approximately true, then past theories cannot be. Past theories are deemed not to be approximately true, because the entities they posited are no longer believed to exist and/or because the laws and mechanisms they postulated aren't part of our current theoretical description of the world. As Laudan put it:
"Because they [most past theories] have been based on what we now believe to be fundamentally mistaken theoretical models and structures, the realist cannot possibly hope to explain the empirical success such theories enjoyed in terms of the truthlikeness of their constituent theoretical claims" (1984a, pp91-92). Then, comes the 'historical gambit': (C) These characteristically false theories were, nonetheless, empirically successful. So, empirical success isn't connected with truthlikeness and truthlikeness cannot explain success. The conclusion is that the realists' potential warrant for (A) is defeated. The 'pessimistic induction' "calls into question the realist's warrant for assuming that today's theories, including even those which have passed an impressive array of tests, can thereby warrantedly be taken to be (in Sellars' apt image) 'cutting the world at its joints'" (Laudan, 1984b, p157).

Clearly, the above argument doesn't show that it is inconsistent for realists to claim that current successful theories are approximately true, even if their predecessors have been false. But realists have to give Laudan his dues. Laudan's reductio, although not logically conclusive, is powerful enough to undermine the explanatory connection between success and approximate truth.

One way to try block this reductio is to weaken premiss (C) above, by reducing the size of Laudan's list. The realist claim is that the meta-inductive basis is not long enough to warrant the pessimistic conclusion, anyway (cf. McMullin 1984, p17; Devitt 1984, pp161-

162). Realists use the following two ways to reduce the basis for meta-induction. On the one hand, they dispute the claim that all theories in Laudan's list were successful. Laudan suggests that a theory is successful "so long as it has worked reasonably well, that is, so long as it has functioned in a variety of explanatory contexts, has led to several confirmed predictions, and has been of broad explanatory scope" (1984a, p110). To be sure, he thinks that this is precisely the sense in which realists claim scientific theories to be successful when they propose the 'no miracle' argument. However, realists normally argue that the notion of empirical success should be more rigorous than simply having the right sort of observational consequences, or telling a story that fits the facts. For any theoretical framework (and for that matter, any wild speculation) can be made to fit the facts—and hence to be successful—by simply writing the right kind of empirical consequences into it. As **John Worrall** has repeatedly stressed, the notion of empirical success that realists are happy with involves the generation of novel predictions which are in principle testable. But if this more rigorous notion of success is adopted, then it is not at all clear that all theories in Laudan's list were genuinely successful. It is doubtful, for instance, that the contact-action gravitational ether theories of LeSage and Hartley, the crystalline spheres theory and the theory of circular inertia enjoyed any genuine success (cf. Worrall 1994, p335; McMullin 1987, p70). Hence they should drop out of Laudan's list.

On the other hand, realists suggest that not all past theoretical conceptualisations of several domains of inquiry should be taken seriously. They argue that only mature theories are at issue, ie, theories which have passed the "take-off point" (Boyd) of a specific discipline. This 'take-off point' is characterised by the presence of a body of well-entrenched background beliefs about the domain of inquiry which, in effect, delineate the boundaries of this domain, inform theoretical research and constrain the proposal of theories and hypotheses. This corpus of beliefs gives a broad identity to the discipline by being, normally, the common ground that rival theories of the phenomena under investigation share. It is an empirical issue to find out when a discipline reaches the 'take-off point', but for most disciplines there is such a period of maturity. (For instance, in the case of heat phenomena, the period of theoretical maturity is reached when such background beliefs as the principle of impossibility of perpetual motion, the principle that heat flows only from a warm to a cold body and the laws of Newtonian Mechanics had become well-entrenched.) But if this requirement of maturity is taken into account, then theories such as the 'humoral theory of medicine' or the 'effluvial theory of static electricity' drop out of Laudan's list. The realist point then is that if we restrict our attention to past mature and genuine successful theories, then premiss (C) above is considerably weakened: if we restrict the meta-inductive basis, it no longer warrants the conclusion that genuine success and approximate truth are dissociated. The 'historical gambit' is neutralised.

Notice, however, that this first move alone is not enough to defeat the ‘pessimistic induction’. For, although it is correct that the list of past theories that realists should worry about isn’t as long as Laudan suggests, it is still the case that at least some past theories that pass all realist tests of maturity and genuine success are still considered false and have been abandoned. The relevant examples are the caloric theory of heat and the nineteenth-century optical ether theories. These theories were both distinctively successful and mature. (For a detailed study of the Caloric theory of heat vis-à-vis the pessimistic induction, you can see my article in SHPS.) If these theories are typically false and cannot be defended as approximately true, then the realist’s intended explanatory connection between empirical success and approximate truth is still undermined. So, realists need to reconcile the historical fact that these mature and successful theories were shown to be false and got abandoned with some form or another of realism. Can they do that?

The general way that realists follow is to try to block premise (B) of Laudan’s reductio, viz., if we hold current theories to be approximately true, then past theories are bound not to be since they posited entities that are no longer believed to exist and laws and theoretical mechanisms that have now been abandoned. Clearly, without this premiss the pessimistic conclusion doesn’t follow. Realists suggest that they have a rather plausible way to refute this premiss. They argue that this premiss is true only if past theories are inconsistent with current ones. But, although taken as a whole most past theories are inconsistent with current ones, many of their theoretical assertions are consistent with what we now believe. In fact, many of their theoretical assertions have been retained in subsequent theories of the same domain. In other words, realists argue that when a theory is abandoned, it’s not the case that the theoretical mechanisms and laws it posited are rejected en bloc. Some of those theoretical claims are abandoned, but surely some others get retained as essential elements of subsequent theories. If this retention is considerable and if the theoretical assertions that were retained were doing most of the work in deriving the successes of past theories, then it’s clearly not the case that these past theories were typically false.

If realists are right, then they can now defend an explanatory link between genuine empirical success and approximate truth. But we must not lose sight of the fact that they have given something up. The ‘no miracle’ argument has been considerably qualified. Realists now argue as follows: Laudan has shown us something important: on pain of being at odds with the historical record, the empirical success of a theory cannot issue an unqualified warrant for the truthlikeness of everything that theory says. Yet, it would be equally implausible to think that, despite its genuine success, everything that the theory says is wrong. The right claim seems to be that the genuine empirical success of a theory does make it reasonable to believe that the theory has truthlike theoretical claims. That is, some but not all of the theoretical claims of the theory have "latched onto" the world.

A discussion of the details of the realist argument would lead us very far afield into current lively debates. But the general line of the realist strategy is to characterise what kinds of theoretical claims are abandoned as false and what are retained. Realists suggest that we should look into the structure and content of past theories and try
to differentiate between the theoretical claims that somehow essentially or ineliminably contributed to the generation of successes and those that were 'idle' components that had no contribution to the theory's success. Then, they attempt to show that it was the components that contributed to the successes of the theory that got retained in the successor theory, whereas the claims that got abandoned were the 'idle' components. The underlying thought is that the empirical successes of a theory doesn't indiscriminably support all theoretical claims of the theory, but rather it is differentially distributed among them. Some of the theoretical assertions, eg, those that are essentially used in the derivation of a prediction, get more support than others. **Philip Kitcher** has recently drawn a relevant distinction between "working posits" and "presuppositional posits" (1993, p149). The interested reader should take a look at Kitcher's book The Advancement of Science, pp140-149; and at my piece 'A Philosophical Study ...', especially, pp178-183.

17. Structural Realism

What is worth looking at in some detail is **Worrall's** answer to the pessimistic induction. This is a species of the general strategy outlined in the last few paragraphs but with some very interesting variations. In a nutshell, Worrall too tries to identify the parts of past abandoned theories that get retained in theory-change, but he thinks that these parts relate to the mathematical structure of a theory rather than to its theoretical content. Worrall suggests that both of the arguments we've been discussing so far are, in a sense, right. The empirical success of mature scientific theories suggests that they have somehow "latched onto" the world, as the 'no miracle' argument suggests. Yet, there is also substantial discontinuity at the theoretical or deep-structural level, as the pessimistic induction suggests. These arguments, however, can be reconciled in a way that defends some form of realism, what Worrall calls **Structural Realism**. Structural realism admits that there is radical discontinuity at the theoretical level of scientific theories, ie, at the level of the description of unobservable entities. Yet, it also recognises a substantial retention at the mathematical-structural level (a level in between empirical laws and theoretical accounts of mechanisms and causes). The suggestion then is that this retention marks an important non-empirical continuity in science, while high-level theoretical accounts of unobservable

entities and mechanisms change radically. But notice here that structural realism is not the mere recording of the (undeniable) fact that there is a continuity at the mathematical level in theory-change. (This is not enough of a philosophical position. Not to mention that this fact can be equally accommodated within a full-blown realist position and an instrumentalist one.) Rather, structural realism is best understood as issuing a new constraint on what can be known and on what scientific theories can reveal. In opposition to scientific realism, structural realism somehow restricts the cognitive content of scientific theories to their mathematical structure together with their empirical consequences. But, in opposition to instrumentalism, structural realism suggests that the mathematical structure of a theory reflects the structure of the world (ie, it reflects real relations between unobservables). So, structural realism defends the following: (1) Scientific theories can, at best, reveal the structure of the underlying physical reality by means of their mathematical structure.
(2) Mathematical equations which are retained in theory-change express real relations between objects for which we know nothing more than that they stand in these relations to each other. (3) Different ontologies (and hence different physical contents) may satisfy the same mathematical structure but there are no independent reasons to believe in one of those as the correct one.

The structural realist position has been advanced by **Henri Poincaré** in the beginning of the century. He, in fact, anticipated the argument from the ‘pessimistic induction’. He noted: “The man of the world is struck to see how ephemeral scientific theories are. After some years of prosperity, he sees them successively abandoned; he sees ruins accumulated on ruins; he predicts that the theories in vogue today will in a short time succumb in their turn, and he concludes that they are absolutely in vain. This is what he calls the bankruptcy of science.” But he went on to say: “His scepticism is superficial; he does not understand none of the aim and the role of scientific theories; without this he would understand that ruins can still be good for something”. For Poincaré scientific theories do not reduce to “practical recipes” as pure instrumentalists think. Rather, successful scientific theories can tell us something about the structure of the world. He referred to the transition from Fresnel’s ether theory of light to Maxwell’s electromagnetism and noted that Fresnel’s mathematical equations are carried over to Maxwell’s theory, although their interpretation changes radically. He then pointed out that “(T)hese equations express relations, and if the equations remain true, it is because the relations preserve their reality. They teach us, now as then, that there is such and such a relation between this thing and some other thing; only this something we formerly called motion we now call it electric current. But these appellations were only images substituted for the real objects which Nature will eternally hide from us. The true relations between these real objects are the only reality we can attain to, and the only condition is that the same relations exist between these objects as between the images which we are forced to put in their place" (International Congress of Physics, Paris 1900).

So Poincaré—and after him Worrall—thought that although the nature of things cannot be known, successful scientific theories can still tell us something about the relations that these things stand to one another. For Poincaré the aim of science was not to discover 'the real nature of things'. He thought that "Still things themselves are not what it [i.e. science] can reach as the naive dogmatists think, but only relations between things. Outside of these relations there is no knowable reality" (1902, p28). And Worrall suggested that the structural realist "insists that it is a mistake to think that we can ever 'understand' the nature of the basic furniture of the universe" (1989, p122).

It turns out however that structural realism faces some interesting problems. First of all, Worrall's position relies on an alleged sharp distinction between the structure of a scientific assertion and its content. However, in modern science structure and nature form a continuum: the nature of an entity or mechanism is given via a structural/mathematical account of its properties and relations. But even if one granted such a distinction, it's not obvious that structural realism can be easily defended as a halfway house between realism and instrumentalism. Laudan has in fact anticipated that one "might be content with capturing only the formal mathematical relations" of the superseded theory within its successor (1981, p40). But he rightly dismissed this view as a viable realist answer since it amounts to the response of "'closet' positivists". In order to make his position realist, the structural realist needs to show that mathematical equations represent real relations in the world which are knowable independently of their relata. In particular, he needs to justify the move from the widespread fact that mathematical equations are retained in theory-change to the claim that they signify real relations between physical objects otherwise unknown. Suppose that the argument is a variant of the 'no miracle' argument (what else could it be?). That is, suppose that the structural realist argues that from the vantage point of the successor theory, the best way to understand why the abandoned theory was empirically successful is to suggest that the retained mathematical equations express real relations in the world. In order to make this move successful, the structural realist first needs to show that the mathematical structure of a theory is somehow exclusively responsible for its predictive success. But, that's not true: mathematical equations alone—devoid of their physical content—cannot give rise to any predictions. If one admits that there is substantive (not just formal) retention at the structural/mathematical level, then one should also admit that some physical content is also retained. Such an admission though would undercut the claim that the predictive success vindicates only the mathematical structure of a theory. Here is a brief example: Worrall suggests that Fresnel correctly identified the structure of light-propagation but misidentified the nature of light: Fresnel's equations were retained in Maxwell's theory, whereas Maxwell had a totally different story about the nature of light. But that's too quick. From Maxwell's vantage point, Fresnel had correctly identified a number of properties of light-propagation, including the theoretical mechanism of propagation and the fact that light-propagation satisfies the principle of conservation of energy. In fact, it is in virtue of the fact that Fresnel's theory had "latched onto" these properties of light-propagation that it was able to be predictively successful. The point of this is that if the empirical success of a theory offers any grounds for thinking that some parts of a theory have "latched onto" the world, these parts cannot be just some mathematical equations of the theory but rather some theoretical assertions concerning some substantive properties as well as the law-like behaviour of the entities and mechanisms posited by the theory. (These include but are not exhausted by mathematical equations). Here again though the interested reader should read carefully Worrall's papers (1989), (1994). Some of the foregoing objections are articulated in some detail in my piece 'Is Structural Realism the Best of Both Worlds?'.

18. The Underdetermination of Theories by Evidence

Realists suggest that acceptance of a mature and genuinely successful theory should be identified with the belief that the theory is approximately true. Non-realists, however, suggest that there is a simple argument against the realist thesis: the argument from the underdetermination of theories by evidence (UTE). It goes like this: two theories that are observationally indistinguishable, ie they entail exactly the same observational consequences, are epistemically indistinguishable, too, ie, there are no positive reasons to believe in one rather than the other. Since, however, for any theory that entails the evidence there are incompatible but empirically indistinguishable alternatives, it follows that no theory can be reasonably believed to be (approximately) true. (The core of this idea was stated by **John Stuart Mill**: "An hypothesis (...) is not to be received as probably true because it accounts for all the known phenomena; since this is a condition sometimes fulfilled tolerably well by two conflicting hypotheses; while there are probably many others which are equally possible, but which, for want of anything analogous in our experience, our minds are unfitted to conceive".)

Currently, this argument is employed centrally by **Bas van Fraassen**. He suggests that UTE shows that there are no reasons to believe more in one of a pair of empirically equivalent theoretical descriptions. Van Fraassen suggests that agnosticism is the right attitude towards the theoretical descriptions of the world. Belief in the approximate truth of a theory is never warranted. Instead, we should only accept a theory as empirically adequate. Since two empirically equivalent theories are equally empirically adequate, van Fraassen suggests that we may accept both, but suspend our judgements with respect to their truth. If there is any kind of distinction between the two theories, it can only be based on pragmatic and not epistemic grounds. We can then see how pressing it is for realists to block UTE.

UTE exploits two well-known facts of theory-construction: (1) A given finite segment of observational data does not uniquely entail a hypothesis which accounts for them. (2) There are alternative theoretical formulations which entail the same body of observational data. The first fact relates to the well-known problem of induction. But UTE is not just ordinary inductive scepticism. In fact, its proponents, such as van Fraassen, are not inductive sceptics. Van Fraassen thinks that there is no reason to doubt that induction works reliably vis-à-vis the observable phenomena. He is, however, sceptical about theoretical knowledge in science, ie, knowledge claims that go beyond the empirical phenomena and regularities. So, UTE disputes the realist claim that there are reasonable ampliative (abductive) inferences on the basis of which scientists go beyond the phenomena and form warranted beliefs about unobservable entities and processes. The second fact above relates to the hypothetico-deductive method of confirmation. But UTE is not the mere recording of the claim that more than one theory can have exactly the same observational consequences. UTE intends to establish that none of these theories can be more confirmed than any other: two empirically congruent theories—that is, two theories that entail the same observational consequences—are equally supported by their consequences. Hence, UTE requires that the entailment of the evidence is the only constraint on the confirmation of a theory.

UTE rests on two premisses: First, the Empirical Equivalence Thesis (EET): for any theory T and any body of observational data E there is another theory T' such that T and T' are empirically equivalent with respect to E. (Two or more theories are empirically equivalent iff they entail exactly the same observational consequences.) Second, the Entailment Thesis (ET): the entailment of the evidence is the only constraint on the confirmation of a theory. So, realists need to block at least one of them. Can they do that?

Let’s first concentrate on EET. It is certainly a bold claim, but there is a sense in which its proof is trivial. You can easily see that given any theory T we can construct another theory T' by just adding any statement we like to T, or by just pairwise permuting all the theoretical terms and predicates of T (eg, we permute ‘electron’ with ‘proton’ etc.). We can also create an empirically equivalent theory by taking the Craig(T) of T (cf. last week’s notes). Or, by just admitting the ‘theory’ T^* “All observable phenomena are as if T is true, but T is actually false”. Clearly, T and T^* are logically incompatible but observationally equivalent by construction.

Realists, however, point out that none of the alternatives so far are really serious challengers. T^*, for instance, is not a theory proper; it is just the denial of the claim that there are theoretical entities. But, as we've seen already, the current issue at stake is not the reality of unobservable entities, but rather the correctness of their theoretical descriptions. Hence, the advocates of UTE need to show that there are or can be proper empirically equivalent scientific theories, that is theories that employ theoretical entities but make incompatible claims about them.

Non-realists usually suggest that the so-called Duhem-Quine thesis offers a constructive proof of EET. The Duhem-Quine thesis can be stated as follows:

> All theories entail observational consequences by means of auxiliary assumptions. So, let's take two theories T and T' such that T together with a set of auxiliary assumptions A entail a body of observational consequences E and T' together with another set of auxiliary assumptions A' entail the same body of observational consequences E. Suppose that there is a fresh piece of evidence E' such that T&A entail E' but T' &A is refuted by E'. *It is in principle possible to save T' from refutation*. That is, it is in principle possible to put the blame to the auxiliary assumptions A' and to find another set of auxiliary assumptions A'' such that the theory T' together with A'' entail E'.

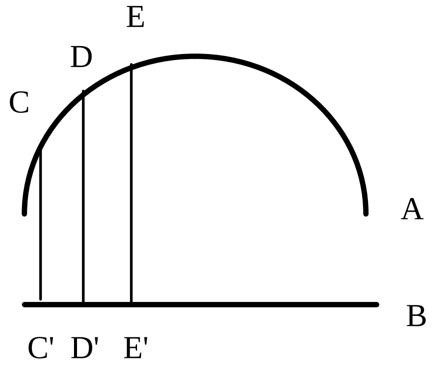
If this thesis is true, then no evidence can discriminate between any two theories T and T'. For, given that theories entail their observational consequences only by means of auxiliary assumptions, if there is a new piece of evidence such that only T—together with a set of auxiliary assumptions A—can accommodate it, the Duhem-Quine thesis asserts that it is always possible that T' can also accommodate it, by suitable adjustments to the relevant auxiliary assumptions. Hence, the Duhem-Quine thesis asserts that *T&A* and *T'&Suitable Auxiliary Assumptions* will be empirically equivalent. In effect, the Duhem-Quine thesis asserts that no evidence can ever tell two theories apart.

One may say a lot of things about the Duhem-Quine thesis. But the most general point is that in all its generality the thesis is trivial. For although it is certainly true that suitable auxiliary assumption are in principle available, it is not at all certain that *non-trivial* auxiliary assumptions can always be found. So, for instance, as a last-ditch attempt to save T' from refutation one may make the recalcitrant evidence E' into an auxiliary assumption. But such a move would be surely trivial and *ad hoc*.

But the Duhem-Quine thesis is a double-edged sword. Suppose that T and T' are already empirically equivalent. The Duhem-Quine thesis suggests that all theories entail predictions with the help of auxiliaries. There is no guarantee that this empirical equivalence will hold when we conjoin the two theories with the relevant auxiliaries so that predictions are generated. That is, suppose that the relevant set of auxiliaries is A. Even if T and T' are empirically equivalent, it is not certain at all that T&A and T' &A will also be empirically equivalent. **Laudan** and **Leplin** generalised this point by arguing that “any determination of the empirical consequence class of a theory must be relativized to a particular state of science. We infer that empirical equivalence itself must be so relativized, and, accordingly, that any finding of empirical equivalence is both contextual and defeasible” (1991, p454). So, the Duhem-Quine thesis cannot offer undisputed support to EET.

But aren’t there some clear cases of empirically indistinguishable theories? A standard example is the following: Let TN stand for Newtonian Mechanics, R be the postulate that the centre of mass of the universe is at rest with respect to absolute space, and V be the postulate that the centre of mass is moving with velocity v relative to absolute space. Then TN & R and TN & V will be empirically indistinguishable given any evidence concerning relative motions of bodies and their absolute accelerations (cf. van Fraassen, 1980, pp46-47). But this is a poor example. For TN & R and TN & V involve exactly the same ontology and ideology for space, time and motion, and hence this is not an interesting case of theoretical underdetermination. Realists can argue that the difference between postulates R and V is immaterial.

Surely, though, so far I’ve been a bit too harsh. For even if realists succeed in diffusing the generality and force of the empirical equivalence thesis (EET), it is certainly true that there are some interesting cases of empirical equivalence. The classic example is that of empirically indistinguishable theories about the geometry of physical space (given by **Poincaré**, 1902, chapters 4 & 5). Here is a sketchy illustration of his example. Suppose that two-dimensional beings inhabit the surface of a hemisphere and they cannot escape from it. (A cross-section of their world is given in A). They try to discover the physical geometry of their world. They use rigid rods to measure distances such as CD and DE and they find them equal.



Soon, they come to the conclusion that they leave on the surface of a sphere. However, an eccentric mathematician of this world suggests that they are collectively mistaken. His hypothesis is that their world is a plane (cross-section B) and not the surface of a sphere (cross-section A). He proposes that there is a universal force, ie, a force that affects everything in this world in the same way. This force makes all moving rods contract as they move away from the centre and towards the periphery. So, he says, the ‘corresponding’ distances C'D' and D'E' are not equal. The measuring rod has contracted upon transportation from D'E' to C'D'. We’ve come to the conclusion that our world is spherical because we have not noticed the deformation of moving rods. How can the inhabitants of this world decide between these two empirically indistinguishable hypotheses? All observable phenomena are as if the shape of the world is the spherical. But the observable phenomena would be exactly the same if the world was flat but a universal force acted on all bodies. This science-fiction story can be easily extended to more realistic cases. In fact, **Hans Reichenbach** (1958, p33 & p66) proved the following general theorem: Suppose we have a physical theory T which suggests that space is curved (eg the General Theory of Relativity). We can construct an empirically indistinguishable theory T' which posits a flat (Euclidean) space, provided that we postulate universal forces which make all moving bodies (eg, moving rods) to contract accordingly. So, roughly, the theories T_1=(Rigid Rods & Curved Space) and T_2=(Universal Forces & Flat Geometry) are observationally indistinguishable. No evidence in terms of coincidences of material bodies and trajectories of moving bodies can distinguish between these theories. Hence even though the strong thesis that for any theory there are interesting empirically indistinguishable alternatives is implausible, a weaker thesis that there are some interesting cases of empirical equivalence is correct.

How do realists react to this last claim? They are generally happy with the existence of some theories that are empirically equivalent. For this fact creates no serious problem for realism. That some domains of inquiry are beyond our ken is not disallowed by realists. More generally, realists are likely to say that the existence of empirically equivalent theories creates a genuine problem only if it is assured that no evidence and no application of any method can possibly discriminate between them. (These are cases of “indefinite underdetermination” (cf. Kitcher, 1992, p97)). But, it is frequently the case that some hitherto empirically congruent theories are told apart by some empirical evidence. For instance, the wave and the corpuscular theories of light were clearly distinguished on empirical grounds by Foucault’s experiment in 1853, concerning the velocity of light in air and in water. If such a resolution becomes available, then no serious case of UTE emerges. Hence, UTE has a bite only if it suggests global scepticism, ie that all hitherto empirically equivalent theories are empirically indistinguishable. But, as we’ve seen, the advocates of UTE have not shown this.

Let’s now turn our attention to the second premiss of UTE, viz., the entailment thesis (ET):

entailment of the evidence is the only constraint on confirmation. This is a crucial premiss. For even if the first premiss (EET) goes through, if ET is shown to be false, then it does not follow that two empirically equivalent theories are necessarily equally supported by the evidence. How can realists block ET?

There are two ways to proceed. First, suppose that one grants that the degree of confirmation of a theory is solely a function of the empirical evidence and its relation to the theory. Still, there is space to deny that the entailment of the evidence is the only constraint on confirmation. Second, one may say that certain ‘super-empirical’ or ‘theoretical virtues’ can influence the degree of confirmation of a theory. Let’s examine these two options in some more detail.

As **Laudan** and **Leplin** (1991) have recently shown one can diffuse the force of UTE by arguing that being an empirical consequence of a hypothesis is neither sufficient nor necessary for being evidentially relevant to a hypothesis. A Hypothesis can be confirmed by evidence that is not a logical consequence of this hypothesis. And conversely, not all logical consequences of a hypothesis are potentially confirming. Let’s take these two in turn. a) Hypotheses can be confirmed by empirical evidence that does not logically follow from them. The typical example here is that the discovery of Brownian motion was widely taken to confirm the atomic theory although it was not a consequence of this. How, generally, can a piece of evidence support a theory without being one of its consequences? Suppose that a piece of evidence E is entailed by a hypothesis H which in turn can be embedded in a more general theory T. Suppose also that T entails another hypothesis H'. E can be said to indirectly support H' although it is not a logical consequence of H'. (So, for example, Brownian motion indirectly supported the atomic theory by directly supporting statistical mechanics.) b) Hypotheses are not confirmed by their empirical consequences. We’ve seen already in connection with the so-called Ravens paradox, that one can consistently deny that positive instances of a hypothesis necessarily confirm the hypothesis. Take for instance the hypothesis that reading of the scripture between the age of 7 and 16 induces puberty in young males. Suppose also that there positive instances of this hypothesis, viz., a number of young males who were forced to read the scripture for 9 years and who have assuredly reached puberty by the age of sixteen. Yet, would we be willing to say that these cases confirm the hypothesis at hand? The answer is clearly negative since we would feel that the evidence is not good and representative enough, it can be easily explained without any loss by a readily available alternative hypothesis etc. The point of this counter example is this: it is precisely because scientists recognise that not all positive instances of a hypothesis are confirming instances that they put additional conditions on the admissibility of evidence, eg. variability, control for spurious correlations etc. How does this line help diffuse UTE? Two theories with exactly the same observational consequences may enjoy differing degrees of evidential support: either because only one of them is indirectly supported by other relevant evidence, or because one of them is not really supported by its positive instances.

Most realists though tend to also adopt the second available option, viz., the claim that certain ‘super-empirical’ or ‘theoretical virtues’ can influence the degree of confirmation of a theory. They suggest that when it comes to assessing scientific theories, we should not just examine them with respect to their empirical adequacy. This is necessary but not enough of its own to make a good theory. We also need to take into account several theoretical virtues such as internal consistency, coherence with other established theories, simplicity, completeness, unifying power, lack of ad hoc features, naturalness. Realists suggest that these values capture, to some extent at least, the explanatory power of a theory and that explanatory power is potentially confirmatory. Take, for instance, two theories T and T' that entail the same body of data $e_1,...,e_n$. Suppose that for every piece of data e_i ($i=1,...,n$) T' introduces an independent explanatory assumption T'_i such that T'_i entails e_i. But suppose T employs fewer hypotheses, and hence unifies the phenomena by reducing the number of independently accepted hypotheses. The claim is that because of this unification T is more confirmed than T'. So, even if two theories are observationally indistinguishable, it’s not clear at all that they have equal explanatory power. If these extra values are taken into account, it won’t be easy at all to find more than one theory that satisfies them to an equal degree. However, non-realists philosophers of science, such as van Fraassen, have objected to this move by denying that explanatory power has anything to do with confirmation and truth. They argue that the foregoing are indeed virtues of theories but they are pragmatic, or aesthetic, rather than epistemic. So realists need to defend the view that these theoretical virtues are truth-tropic. This is still pretty much an open issue and clearly relates to the credentials of Inference to the Best Explanation. **Ernan McMullin**, for instance, suggests that, in effect, explanation (as well as predictive accuracy) are the constitutive aims of science and it is only rational to choose the theory with the most explanatory power. He adds that the theoretical virtues are those that scientists use to characterise a “good theory” and those that have been traditionally “thought to be symptoms of truth generally” (1987, p66-67). As he put it “The values that constitute a theory as ‘best explanation’ are, it would seem, similar to those that would qualify a statement as ‘true’”. This move somehow identifies ‘best explanation’ with ‘truth’. Hence, although it makes the choice of the best explanation the rational thing to do, it delivers the realist aim of truth too easily. So, I think, realists need to do more work here. One way to proceed is to try to show that the history of science itself suggests that the best way to view these theoretical virtues is as epistemic. (This is roughly the line followed by

Richard Boyd.) Another promising line may be to suggest that even non-realists a là van Fraassen need to treat explanatory power as potentially confirmatory, if they are to have grounded judgements of empirical adequacy, ie if they want to avoid being total sceptics. (I develop this line in my piece ‘On van Fraassen’s Critique...’.)

The two foregoing options available to realists have in fact been met in **Clark Glymour**’s theory of confirmation. Glymour’s important contribution to the debate is two fold: (1) he showed that even if two theories entail the same observational consequences, there are still ways to show that they may be differentially supported by them; (2) he argued that theoretical virtues, especially the explanatory power and unifying power bear on the confirmation of a theory. As he put it: “Confirmation and explanation are closely linked (...)” (1980, p376). Here is a simple example as to how this can block UTE (or at least some cases of it). Suppose that a fundamental quantity F in a scientific theory T is analysed as the sum of two primitive quantities F_1 and F_2 (ie, $F=F_1+F_2$). One can then construct a theory T' which employs F_1 and F_2 instead of F and is clearly empirically equivalent to T, (T' has precisely the same observational consequences as T). Are T and T' equally supported by the evidence? Glymour suggests that there must be a scientifically significant reason for breaking up F into two components. If there is such a reason, then it must be supported by an independent argument as to how the individual values of F_1 and F_2 can possibly be computed. Hence, there will be an excess content in T' (ie, the independently specifiable values of F_1 and F_2) which can be used to differentiate it from T, ie, to suggest some possible tests for the values of F_1 and F_2. If, on the contrary, there is no scientifically significant reason for introducing F_1 and F_2, and if there is no way to calculate their individual values, then one can say that T is better supported than T'. For T' has more untested hypotheses than T, while T is more unifying than T' because it minimises the number of independently acceptable hypotheses (1980, pp356-357). In effect, Reichenbach’s recipe for creating empirically equivalent space-time theories (cf. a few paragraphs above) is an instance of the previous ad hoc technique (cf. Glymour, 1980, p365). But if a piece of evidence can support T more than T', although both T and T' entail it, there are good reasons to believe in T more than T': the existence of empirically equivalent theories does not yield epistemic indistinguishability.

To sum up, UTE is not as powerful an argument as it seems. Realists can succeed in blocking it. At worse, there is a stand-off between realists and agnostics which reflect their differences about the status and worth of theoretical virtues in science.

19. van Fraassen’s Constructive Empiricism

So far, we’ve been citing van Fraassen’s objections to realism without discussing his own positive alternative. We’ve already seen that van Fraassen is an agnostic concerning the theoretical claims made by science. He doesn’t believe that they are outright false; but he doesn’t believe that they are (approximately) true either. He defends a position called ‘constructive empiricism’. According to this: (1) the aim of science is to produce empirically adequate theories, ie theories that save the phenomena; and (2) acceptance of theories need not involve the belief that they are true, but rather only the belief that they are empirically adequate.

Van Fraassen’s critique of realism stems largely from his defence of an empiricist view of science, in particular a view that aims to avoid unnecessary metaphysical commitments. The thrust of his position is that belief in the truth of the theoretical claims of a theory is “supererogatory” because one can just believe in the empirical adequacy of the theory and do exactly as well. He suggests that “we can have evidence for the truth of the theory only via evidential support for its empirical adequacy”. Hence we can never have more reasons to believe in the truth of a theory than to believe in its empirical adequacy. (In fact, since the truth of the theory entails its empirical adequacy, one can use a well-known theorem of the good old probability calculus, to show that the probability that a theory is true is less than or equal to the probability that it is empirically adequate; (if T entails E, then Prob(T) is less than or equal to Prob(E)). So, van Fraassen suggests that, in line with the anti-metaphysical Occam’s razor, belief in the truth of the theory is redundant.

Constructive empiricism is supposed to be a kind of intermediary position between old empiricist-instrumentalist positions and scientific realism. Unlike traditional instrumentalists, van Fraassen agrees with realists that theoretical (ontological and explanatory) commitments in science are ineliminable. (That is, he agrees with realist position (1) as specified in section 12 of lecture notes III.) Van Fraassen accepts that scientific theories should be taken at face-value and be understood literally as trying to describe the reality behind the phenomena. So, theoretical terms like ‘electron’ and ‘proton’ shouldn’t be understood as useful ‘shorthands’ for complicated connections between observables. Rather, they should be taken to refer to unobservable entities in the world. Scientific theories are then taken as constructions with truth-values: true, when the world is the way the theory says it is, false otherwise (cf. 1980, p10 & p38). So, van Fraassen’s position is very different from semantic and Duhemian instrumentalisms.

Similarly, van Fraassen explicitly denounces the Carnapian two-language view of scientific theories (cf. section 5 of Lecture Notes II). Instead, he admits that all observation is theory-laden. For van Fraassen “all our language is thoroughly theory-infected” (1980, p14), in fact, theory-infected to the extent that if we started to chuck out the so-called theoretical terms, then “we would end up with nothing useful” (ibid.). Yet, he remains an empiricist, albeit a modern-day one, in that he still believes that “our opinion about the limits of perception should play a role in arriving at our epistemic attitudes towards science” (1985, p258). How can he marry empiricism with the claim that all language is theory-infected?

Van Fraassen assigns a privileged epistemic role to observability, but he dissociates it from the old empiricist demand of a description in an observational vocabulary. This is a key move. He holds on to a distinction between observable and unobservable entities, which he thinks can play an important epistemic role, but he denies that this distinction mirrors the (old empiricist) line that separated theoretical from observational terms (cf. 1980, p14 & p54). He thinks that observability should guide (in fact, determine) what we believe, but the description of our belief-content can be given in a thoroughly theory-infected language. Doesn’t this view sound odd?

Well, opinions differ here. But van Fraassen seems to be consistent. For him an entity is observable if a suitably placed observer can perceive it by the unaided senses (cf. 1980, p16). So, Jupiter’s satellites are currently observed only through telescopes, but an astronaut launched near Jupiter could observe them with naked eye. This, he thinks, is an empirical distinction: it is science, in particular our current scientific theories, that delineate what entities are observable and what go beyond the limits of observability. He says: “To delineate what is observable, however, we must look to science—and possibly to that same theory—for that is also an empirical question” (op.cit., p58). In particular, the limits of observability are characterised by empirically discoverable facts about “us qua organisms in the world, and these facts may include facts about the psychological states that involve contemplation of theories” (ibid.). So, whether an entity is or is not observable is only to do with physiological and psychological facts about humans, in particular, with the contingent fact that some entities may be too little to be seen by the naked eye. The description of an observable entity is, however, a different matter. One can describe a table using the terms of modern science as “an aggregate of interacting electrons, protons, and neutrons”. Yet, a table is an observable thing. One can believe in the existence of the table, use theoretical language to describe what a table is and yet remain agnostic about the correctness of its current description by science. Van Fraassen can consistently make all these claims, but precisely because he is an agnostic and not an atheist. He doesn’t want to say that the entities posited by science do not exist. If he did say so, he would be in all sorts of troubles if he wanted to stick to the idea that all language is theory-infected (try to think why). What he says is that the fact that he uses the current descriptions to characterise observable objects doesn’t commit him to the correctness of these descriptions.

Be that as it may, van Fraassen suggests that the observable/unobservable distinction draws the borders between what is epistemically accessible and what is not: all statements about the unobservable world are undecidable, in that no evidence can warrant belief in theoretical claims about the unobservable world. There is no doubt that some things are visible to the naked eye, whereas others are not. (Although it is debatable whether only things that are visible to the naked eye should count as observable. Many philosophers, most notably Ian Hacking, have persuasively suggested that van Fraassen's notion of observability is too narrow and that instrument-aided observations are no less reliable than eye-based one. This line can be developed into a different argument for realism: entities that are not visible to the naked eye can nonetheless be observed through instruments. The interested reader should read Hacking's 'Do we See Through a Microscope?'). The question here is why and whether this fact should have any special epistemic significance. Van Fraassen does not just want to point out the interesting fact that some things are unobservable while some others are not. His distinction between observables and unobservables is intended to play an epistemic role. But why should it be the case that unobservability is tantamount to epistemic inaccessibility and observability is tantamount to epistemic accessibility?

The idea is that unaided senses can decide claims about observables but they cannot decide claims about unobservables. But unaided senses alone, ie, senses without the aid of instruments, can decide nothing but a tiny fraction of the things that scientists claim they know. Senses alone cannot decide even elementary processes such as measuring temperatures, apart from the crude sense that a body with high temperature 'feels' hot. Van Fraassen normally argues that as far as observable phenomena are concerned, it is always possible that a human being can be in such a position that she can decide claims about them by means of unaided senses. For instance, Jupiter's satellites are observed through a telescope but an astronaut launched to Jupiter can observe them by naked eye. The trouble with this suggestion is that it is generally incorrect. Unaided senses cannot even decide claims concerning the empirical adequacy of theories. For instance, our theories say that the temperature in Pluto is extremely low. But nobody can check this—even in the crude sense of 'feeling cold'—by being transported to Pluto, the reason being that no human being can possibly survive these low temperatures. So, a recourse to unaided senses is too poor a move to sustain the alleged epistemic relevance of the observable/unobservable distinction. This aspect of van Fraassen's scepticism has been extensively discussed and heavily criticised by many philosophers of science. The interested reader should take a look at **Paul Churchland**'s (1985) and **Wesley Salmon**'s (1985). (Salmon has argued that empiricism is consistent with the belief in unobservables and cites Reichenbach as an example of this.)

Generally, what realists suggest is that it is one thing to demand extra caution in knowledge-claims about the unobservable world, especially in the light of the fact that scientists have been in error in some of their beliefs about it; but it is quite another to adopt an implausible account of knowledge, which excludes from knowledge any claim that goes beyond what can be observed with naked eye, felt and the like. Sceptic philosophers of science are right to point out the need for caution, but wrong insofar as their demand for caution leads them to banning any knowledge of the unobservable world.

So far, we only criticised the alleged epistemic significance of the distinction between observable and unobservables. Van Fraassen has other arguments against realism. Most of the debate, though, revolves around the status of theoretical virtues: are they pragmatic (van Fraassen) or are they epistemic (realists). But there is something else worth saying. As we saw in the beginning of this section, van Fraassen's central thought is that realism is "inflationist" because belief in truth is logically stronger (and hence less probable) than belief in empirical adequacy. But realists can answer to this point as follows: the fact that the probability of the theory being true is less than or equal to the probability that it is empirically adequate does not entail that the probability that the theory is true is not high, or at any rate, high enough to warrant the belief that the theory is true. Surely, realists stick their necks out much further than constructive empiricists, but this is after all the price for trying to push back the frontiers of ignorance and error. (Here again, more details on these matters are given in my piece 'On van Fraassen's critique...'.)

20. Fine's Natural Ontological Attitude

Arthur Fine has recently suggested that both instrumentalism and realism are, for different reasons, in the wrong. They are both "unnatural attitudes" to science, more extraneous attachments to the body of science than natural garbs of this body (cf. 1986). But then what is the natural attitude to science? Fine developed his own 'third way', what he called Natural Ontological Attitude (NOA).

Fine suggests that both realism and instrumentalism in all of its guises are infationary and "hermeneutic" philosophical attitudes. Scientific realism aspires for an "outside" authentication of science: that science is about the world. Instrumentalists and other anti-realists, on the other hand, aspire for an "inward" authentication of science: science is about us humans and our relations with the observable world. To these views, Fine opposes his own deflationism. NOA, briefly, suggests: "Try to take science on its own terms, and try not to read things into science" (1986, p149).

One of the most central features of NOA is that, according to Fine, it does not inflate the concept of "truth". Rather, NOA recognises in "truth" a concept "already in use" in science (as well as in everyday life) and abides by "the standard rules of usage" (1986, p133). These rules involve the usual "Davidsonian-Tarskian referential semantics" (ibid.). [Roughly, the semantic captured in the schema ''Snow is white' is true iff snow is white'. Once more, for more on this look at the appendix of Boyd's piece 'Confirmation, Semantic and the Interpretation of Scientific Theories'.)] Fine thinks that ascriptions of truth in general and scientific truth in particular are to be understood "in the usual referential way so that a sentence (or statement) is true just in case the entities referred to stand in the referred to relation" (1986, p130). Besides, as a consequence of abiding by the standard truth-semantics, NOAers believe that a theory issues specific ontological commitments to "the existence of the individuals, properties, relations, processes and so forth referred to by the scientific statements that we accept as true" (1986, p130). In particular, NOAers are happy to be committed to the existence of unobservable entities, if the presumed true scientific statements involve reference to unobservables. As for the epistemic attitude towards scientific claims, NOAers agree that the degree of confidence in the truth of a given scientific theory will determine the degree of belief in the existence of the entities posited by this theory, where the former is "tutored by ordinary relations of confirmation and evidential support, subject to the usual scientific canons" (1986, p130).

One may wonder here whether scientific realists would not subscribe to all this. I think that, with a few provisos, they would and they should. But, first, it may be interesting to see why Fine thinks that scientific realists inflate the concept of truth already in use in science.

Fine's contention is that realists (as well as anti-realists of all sorts) accept what he calls the "core position" concerning truth, viz., the results of scientific investigations are true, if at all, on a par with more homely truths. Yet, he says, realists "add onto" this core position by saying that these truths are made true by and are about the world (while anti-realists "add onto" this core position by arguing that they are true because the right kind of epistemic condition obtains) (cf. 1986, pp128-129). Many philosophers rightly think, however, strictly speaking, there is no such thing as a neutral core position that realists and anti-realists happily share in common (cf. Musgrave, 1989, p384). To say that a statement is true, be it about 'homely' tables and chairs or about 'exotic' electrons and monopoles, is to offer an answer to the question 'in virtue of what is it true?'. And, as it happens, different philosophical schools offer different answers to this question, even though they may agree on the Tarskian disquotational schema as the basic 'grammar' of the predicate 'is true in L', and even though they may (partly) agree on the final product, i.e. on what claims can be said to be true. (Here again, we're getting into deep waters. Suffice it to say that most realists think of truth as a non-epistemic notion, ie, they think that ascriptions of truth are independent of our capacities to verify of justify them, whereas most anti-realists (but not sceptics a là van Fraassen) think that truth is an epistemic concept, ie, it is conceptually linked with some epistemic notions such as verfication, ideal justification, warranted assertibility etc.) Be that as it may, Fine suggests that in doing science and in reflecting on aspects of the scientific game there is no need to add onto the core position as he characterised it. All that is required in order to account for the ontological commitments of science and the regulation of scientific belief is to "maintain parity" between ascriptions of homely truths and ascriptions of scientific truths. Then, if one admits that homely truths are simply true, it follows that if some scientific assertions are true at all, they are simply true; not true of the World but true simpliciter; true, "on a par with more homely truths" (1986, p128). Instead of doing just that, Fine says, realists give a special interpretation to the concept of truth as used in science. They adopt a "desk-thumping, foot-stamping" attitude towards scientific truths. In an attempt to explain the robust sense which they attribute to their truth-claims, they argue: "There really are electrons, really!" (1986, p129). These claims are said to be about reality—"what is really, really the case" (ibid.).

Most realists are likely to say that for them all ascriptions of truth, be they about 'homely' or 'exotic' beliefs, have the following kind of robust gloss put on them: they are non-epistemic claims in that they do not reduce the concept of truth to primarily epistemic notions such as 'conclusive verification', or 'warrantedly assertibility', or 'idealised justification'. Othewise realists, as Fine demanded, would say that they maintain parity between ascriptions of homely truths and ascriptions of scientific truths. Why, and to what extent, does this understanding of truth go beyond the conception of truth as used in science and that NOAers are happy with?

One should note here that Fine is, in a way, right in suggesting that the question whether, say, electrons are real does not arise naturally. Yet, interestingly enough, the question has arisen and it was not the defenders of realism that raised it. My own view is that the apparent force of Fine's claim that realists inflate truth rests on not taking on board the development of the debates over realism in science. Scientific realists have always been very keen to take scientists' claims about unobservable entities, processes and mechanisms at face value. As we saw in the previous lecture, their main point has always been that the world is populated by the unobservable natural kinds posited by well-confirmed scientific theories and that well-confirmed scientific beliefs are true of the entities they posit. Most current realists would agree with Fine that ontological commitments in science are just an extension of the ontological commitments to common-sense objects (cf. Devitt, 1984;

Newton-Smith, 1989). As we saw in the previous lectures, the realist views have been challenged by instrumentalists of all sorts. Scientific unobservables had been seen as abbreviatory schemes for the description of the complex relationships between observables, auxiliary devices, façons de parler etc. Similarly, theoretical claims have been seen as claims about the actual and possible behaviour of observable entities; instrumental steps; reducible to sets of observational statements and the like. It was among this set of re-interpretations of claims about unobservables that, for instance, (as we saw already in section 5, lecture notes II) Herbert Feigl suggested the distinction between “the factual reference” of theoretical terms and their “epistemic reduction”, ie, the evidential basis on which we may assert their presence, and talked about “the independent existence of the referents of hypothetical constructs” (1950, pp48-49). The claim of independent existence was called forth precisely in order to ground the view that assertions about unobservables should not be equated with the evidence scientists may have for their presence, and hence the view that assertions concerning the truth of theoretical statements should not be reduced to claims about connections between observables.

So, it was mainly because non-realists of all sorts either tried to challenge the existence of such entities or attempted to tie existence to whatever can be epistemically accessed, that realists sometimes had to follow the foot-stamping attitude, exemplified in ‘electrons really exist’. As Smart (1963, p35) pointed out, the ‘real’ or ‘really’ that realists appealed to were precisely meant to block equating scientific unobservables with theoretical fictions (like lines of force), logical constructions (like the average height), or non-existent objects (like unicorns). But, insofar as theoretical assertions are taken at face value and are understood literally, realists need only assert that if the theory is true, then what it says about its domain obtains; and conversely, the theory is true only if its domain is as the theory describes it. Hence, insofar as NOA treats scientific theories as truth-valued constructions that issue specific ontological commitments to unobservables, then NOAres and realists are in the same boat.

Study Questions

1. Outline Boyd’s argument for scientific realism. Do you think it’s a good argument? What are the main objections against it?

2. “The success of science is not a miracle. It is not even surprising to the scientific (Darwinist) mind. For any scientific theory is born into a life of fierce competition, a jungle red in tooth and claw. Only the successful theories survive—the ones which in fact latched on to actual regularities in nature”. Discuss.

3. What is the argument from the ‘pessimistic induction’? Does it destroy the ‘no miracle’ argument?

4. How does Structural Realism try to block the ‘pessimistic induction’? Is it successful?

5. State clearly the argument from the underdetermination of theories by evidence. Are two empirically equivalent theories necessarily equally confirmed by the evidence?

6. Critically assess van Fraassen’s use of the distinction between observable and unobservable entities.

References

Boyd, R. (1984) ‘The Current Status of the Realism Debate’, in J. Leplin (ed.) Scientific Realism, University of California Press.

Churchland, P. M. (1985) ‘The Ontological Status of Observables: In Praise of Superempirical Virtues’ in P. M. Churchland & C. A. Hooker (eds.) Images of Science, The University of Chicago Press.

Devitt, M. (1984) Realism and Truth, (Second Revised edition 1991), Oxford: Blackwell.

Fine, A. (1986) The Shaky Game, Chicago IL: The University of Chicago Press.

Fine, A. (1986a) ‘Unnatural Attitudes: Realist and Instrumentalist Attachments to Science’, Mind, **95**, pp.149-179.

Fine, A. (1991) ‘Piecemeal Realism’, Philosophical Studies, **61**, pp.79-96.

Glymour, C. (1980) Theory and Evidence, Princeton: Princeton University Press.

Hesse, M. B. (1976) ‘Truth and Growth of Knowledge’ in F Suppe & P D Asquith (eds.) PSA 1976 Vol.2, pp.261-280, East Lansing MI: Philosophy of Science Association.

Kitcher, P. (1993) The Advancement of Science, Oxford University Press.

Laudan, L. (1981), ‘A Confutation of Convergent Realism’, Philosophy of Science **48**, pp. 19-49.

Laudan, L. (1984) ‘Explaining the Success of Science’, in J Cushing et al. (eds.) Science and Reality, Notre Dame University Press.

Laudan, L. (1984a) Science and Values, University of California Press.

Laudan, L. (1984b) ‘Discussion: Realism without the Real’, Philosophy of Science, **51**, pp.156-162.

Laudan, L. & Leplin, J. (1991) ‘Empirical Equivalence and Underdetermination’, Journal of Philosophy, **88**, pp.449-472.

Lipton, P. (1991) Inference to the Best Explanation, Routledge and Kegan Paul.

McMullin, E. (1984) ‘A Case for Scientific Realism’ in J Leplin (ed.) Scientific Realism, University of California Press.

McMullin, E. (1987) ‘Explanatory Success and the Truth of Theory’, in N Rescher (ed.) Scientific Inquiry in Philosophical Perspective, University Press of America.

Musgrave, A. (1988) ‘The Ultimate Argument for Scientific Realism’, in R Nola (ed.) Relativism and Realism in Sciences, Dordrecht: Kluwer Academic Press.

Newton-Smith, W. H. (1989a) ‘Modest Realism’ in A Fine & J Leplin (eds.) PSA 1988 , Vol.2, pp.179-189, East Lansing MI: Philosophy of Science Association.

Papineau, D. (1993) Philosophical Naturalism, Oxford: Blackwell.

Psillos, S. (1994), ‘A Philosophical Study of the Transition from the Caloric Theory of Heat to Thermodynamics: Resisting the Pessimistic Meta-Induction’, Studies in History and Philosophy of Science 25: 159-190.

Psillos, S. (1995), ‘Is Structural Realism the Best of Both Worlds?’ Dialectica 49: 15-46.

Psillos, S. (1996), ‘On van Frrassen’s Critique of Abductive Reasoning’, The Philosophical Quarterly 46: 31-47.

Poincaré, H. (1902) Science and Hypothesis, New York: Dover Inc.

Reichenbach, H. (1958) The Philosophy of Space and Time, New York: Dover Publications.

Salmon, W. (1985) ‘Empiricism: The Key Question’ in N Rescher (ed.) The Heritage of Logical Positivism, University Press of America.

Worrall, J. (1989), ‘Structural Realism: the Best of Both Worlds?’, Dialectica 43, pp.99-124.

Worrall, J. (1994), ‘How to Remain Reasonably Optimistic: Scientific Realism and the ‘Luminiferous Ether’’, in D. Hull et al. (eds) PSA 1994, Vol.1. East Lansing: Philosophy of Science Association, pp.334-342.

Van Fraassen, B. (1980) The Scientific Image, Oxford: Clarendon Press.

Van Fraassen, B. (1985) ‘Empiricism in Philosophy of Science’ in P. M. Churchland & C. A. Hooker (eds.) Images of Science, The University of Chicago Press.