ΦΙΛΟΣΟΦΙΑ ΤΗΣ ΓΛΩΣΣΑΣ

ΣΗΜΕΙΩΣΕΙΣ

ΣΤΑΘΗΣ ΨΥΛΛΟΣ

ΑΘΗΝΑ 2001

0. ΠΡΟΛΟΓΟΣ

Οι σημειώσεις που έπονται καλύπτουν ένα βασικό μάθημα φιλοσοφίας της γλώσσας. Για λόγους ευκολίας έχουν γραφεί στα αγγλικά. Θα επιδιωχθεί όμως η επόμενη έκδοση να είναι στα ελληνικά. Ο συγγραφέας ζητά την κατανόησή σας.

Αθήνα 21/01/02

Σ. Ψ.

Βασική Βιβλιογραφία

Εγχειρίδια

William Lycan: Philosophy of Language: A Contemporary Introduction, Routledge, 1999.

Alexander Miller: Philosophy of Language, 1998, UCL Press.

Michael Devitt & Kim Sterelny: Language and Reality: An Introduction to the Philosophy of Language, 1987, Blackwell.

Willard v. O. Quine: Philosophy of Logic, 1970, Harvard U.P.—Ελληνική Μετάφραση από τις εκδόσεις Δαίδαλος-Ζαχαρόπουλος.

Ian Hacking: Why Does Language Matter to Philosophy?, 1975, Cambridge UP.

Crispin Wright & Bob Hale: A Companion to the Philosophy of Language, Blackwell.

Πολλά σημαντικά άρθρα μπορούν να βρεθούν σε διάφορες ανθολογίες.

Peter Ludlow (ed.) Readings in the Philosophy of Language, 1997 MIT Press.

Για το πρόγραμμα του Frege

Frege: The Foundations of Arithmetic, Northwestern University Press.--Greek Translation available.

Frege: Philosophical Writings of Gottlob Frege, 1970, Blackwell.

Ελληνική Μετάφραση του "On Sense and Reference" στο περιοδικό Δευκαλίων, τ. 17..

Michael Beany: Frege: Making Sense, 1996, Duckworth.

Michael Beany: The Frege Reader, 1997, Blackwell.

Joan Weiner: Frege, 1999, Oxford U.P.

Michael Dummett: Frege's Philosophy of Language, Duckworth.

Michael Dummett: "Frege" in Truth and Other Enigmas, 1978, Duckworth.

Michael Dummett: "Frege's Distinction Between Sense and Reference", in Truth and Other Enigmas, 1978, Duckworth.

Για το πρόγραμμα του Russell

Russell: "On Denoting" (in several anthologies) and in B. Russell: Logic and Knowledge.

Russell: "Descriptions" in A. W. Moore (ed.) Meaning and Reference, Oxford U.P.

Russell: The Problems of Philosophy, chapter 5, Oxford University Press.
Strawson: "On Referring" reprinted in A. W. Moore (ed.) Meaning and Reference, Oxford U.P.

I. INTRODUCTORY REMARKS

Γλωσσολογία vs Φιλοσοφία της Γλώσσας
φωνολογία, σύνταξη (γραμματική δομή), σημασιολογία
Θεωρίες σημασιολογίας

Salient features of language

1. System of communication (conveying information) between ourselves and others.
2. More than communication: non-informational or not purely informational social interactions: to greet, question, command, joke, offend, abuse, intimidate, etc.
3. Private uses: we talk to ourselves—we think in language (though contentious that all thinking is linguistic).

Other Features

a. Stimulus Independence: unpredictability of utterances—no direct connection between stimulus and utterance
b. Abstractness: ability to abstract from many details of the situation
c. Arbitrariness: linguistic symbols have no intrinsic necessary connection with their referents (Humpty Dumpty)
d. Medium Independence: too many different media
e. Productivity: knowledge of few words + modes of construction enable us to understand new sentences—unlimited number of sentences
f. Power: it enables us to deal with a host of issues.

What makes language to have all these properties?

N.B. We understand statements, expressions, words. We can understand statements (etc). that we hear for the first time. We can understand a statement/story even if it is false. What exactly does this ability consist in? Descriptively: a) we understand the meaning of individual words; b) we understand something about how they are strung together. But what is it to have this understanding (linguistic competence)?

Property of people/property of mind

But semantic issues can arise without reference to understanding. What is the meaning of the word “snow”? So consider the linguistic expressions in themselves, as objects of study in their own right.

The central philosophical questions:

- What is it for a string of marks or noises to be meaningful?
- What is it in virtue of which a particular string has the distinctive meaning it does?
- How does language acquire semantic content?
- What does understanding of a statement consist in?
- What are meanings?
- How does language hook onto the world?
- What is the vehicle of meaning? (semantic holism vs semantic atomism)
- How does truth relate to meaning?

These are philosophical questions because they are not (necessarily) descriptive. They call for general philosophical theories of semantic content and how this is, if at all, possible.

- **Interplay of syntax and semantics.**

Snow is white
Bob is a better student than Tom
Water is made of oxygen and hydrogen
Honesty is a virtue
Humans are higher animals
Triangularity is a property
Everyone is happy
No-one lives for ever
This sentence is false (or, I am false)
Hooligans behave violently
Green ideas sleep violently

All dada are gaga

Is tall snow because

A fhfur riuiyt tuhhys esr

I. Grammar and meaning

1. grammatical meaningful sentences
2. grammatical but problematic (meaningless?) sentences

e.g. Green ideas sleep violently

3. grammatical but more problematic (meaningless?) sentences

e.g., All dada are gaga

4. ungrammatical sentences

a. Is tall snow because

b. A fhfur riuiyt tuhhys esr

N.B. 2 and 3 have different defects. Important to distinguish between symbol and its interpretation. (Syntax and semantics). Interpretation is one way to think of the semantic value of a symbol.

4a and 4b have different defects.

II. This sentence is false

Meaningful? Yet, if it's true, it's false and if it is false, it is true. (the paradox of the liar)

C: C is false

C iff C is true

Substitute C with what it says.

C is false iff C is true.

Note: You cannot talk about a language in the very same language. Paradoxes.

So, distinction between language and meta-language. The study of a language requires a meta-language. Meta-language has its own "morphology" (syntax). In particular, it has to have names for the sentences and the words of the language. (These names will be words too, but they have to be distinguished from the words of the language.) It must also have expressions (sentences, words) which translate into the meta-language the corresponding expressions of the language.

(homophonic translation—translation nonetheless)

Convention: We use quotation marks " " to state the name of an expression of L in the meta-L. e.g., "George", "Snow is white" etc.

Paradox arose because we didn't respect this distinction between language and meta-language

"C" is true iff C

This is a sentence of the meta-language.

Moral: "is true" is a meta-linguistic predicate.

T: "Snow is white" is true iff snow is white.

T is a sentence in the meta-language. All semantic properties are meta-linguistic.
Is T trivial? Isn't there the same sentence on both sides of the bi-conditional?
NO! In one case the sentence is used, whilst in the other it is mentioned.

- **NOTE on USE and MENTION**

1. John is tall

(1) mentions the person John, but what appears in the sentence is not John but the name of John. (1) mentions John but does not use him. It uses John's name but does not mention it.

Issue: in order to say something about (or mention) anything it is necessary to use a name (or some other means of designation) for it. (In order to talk about quarks, I need to use a name for them.)

Fundamental confusion between a thing and its name—in particular when the subject matter is language itself.

2. John consists of four letters.

(2) is either nonsense, or it is about John's name. But (2) violates the use/mention distinction. For (2) doesn't have a name for John's name, but John's name.

Quotation marks: (Convention) a name of a word is given in quotation marks.

3. "John" consists of four letters.

(3) mentions John's name; it doesn't mention John.

Example:

4. California is a state
5. California has ten letters
6. "California" is a state
7. "California" has ten letters
8. "California" is a name for California
9. ""California"" is a name of a name of California

4, 7, 8, 9 are true, whereas 5,6 are false.

Back to schema T.

"X" is true iff p

"X" is a name for sentence p, whereas p is the sentence. Sentence p is used in the right-hand side, but it is only mentioned in the left-hand side (its name is used).

So, schema T states necessary and sufficient conditions for the truth of the sentence named.

Quine: semantic ascent.

T: "Snow is white" is true iff snow is white.

"Quotation marks make all the difference between talking about words and talking about snow".

III. What is the meaning of an expression X?

Snow is white
Bob is a better student than Tom
Water is made of oxygen and hydrogen
Honesty is a virtue
Humans are higher animals
Triangularity is a property
Everyone is happy
No-one lives for ever

Proper names (Bob), General nouns (snow, water), Abstract nouns (honesty, triangularity), Class terms (humans), pronouns (quantifiers)
Can we have a unified theory of what the meanings of these diverse expressions consist in?
Indexicals: I, here, now etc. What is their meaning?
Types vs tokens

II. THE FREGEAN LEGACY

Gottlob Frege (1848-1925) German mathematician, philosopher and logician.

First attempt to offer a systematic theory of meaning

Central idea: essential to a theory of meaning is the semantic value of an expression.
"Bedeutung" (reference, but in fact more general)

Semantic value: that feature of an expression which determines its contribution to the truth (or falsity) of the sentences in which it occurs.

- Quick reminder of the basic features of the propositional and predicate calculus. The language of modern symbolic logic.

Syntax:

- Specification of a vocabulary
- Set of rules which determine which sequences of expressions constructed from this vocabulary are grammatical.

The syntax of propositional logic
Propositional variables, connectives, parentheses.

The syntax of predicate logic:
The above plus quantifiers and variables. (Why do we need a quantificational logic?)
Some valid arguments cannot be expressed in the propositional calculus.

e.g. All men are mortal
Socrates is a man
Therefore, Socrates is mortal.

N.B. syntax doesn't require knowledge of meanings

The syntax of logic could be done purely in terms of shapes/forms and rules of ordering—rules which say which forms are well-formed formulas.

Fregean syntactic categories: proper names, predicates, declarative sentences, sentential connectives and quantifiers.

Semantic properties and validity

Validity can be expressed in syntactic terms, e.g., arguments of the form 'a&b, therefore b' are valid.

Yet, intuitively, validity has to do with truth-preservation. An argument is valid if whenever the premises are true the conclusion has to be true. So, truth enters the characterisation of validity. This is expressed clearly in the method of truth-tables.
N.B. The truth-tables specify the meaning of the logical connectives.

❑ Semantic value ("Bedeutung")

Definition: The semantic value of any expression is that feature of it that determines whether sentences in which it occurs are true or false.

The beginnings of a semantic theory: an assignment of a semantic property (truth or falsity) to the sentences of a language which determines the validity of the inferences in which those sentences appear as constituents.
Why build a theory of meaning around truth-linked concepts?

- Communication—conveying information
- Translation
- Inference

In all these, truth features prominently. E.g., can two expressions be translations of each other and yet one be true and the other false?

Thesis 1: The semantic value of a sentence is its truth-value (truth or falsity).

This is motivated by the study of validity in arguments.

Thesis 2: (Compositionality) The semantic value of a complex expression is determined by the semantic values of its parts.

Propositional logic: atomic vs molecular sentences.

- Atomic sentences are assigned truth-values. (Not further analysable)
- Molecular sentences: their truth value is a function of the truth-values of their constituents, given the connectives.

Predicate logic: atomic vs molecular sentences.

- Atomic sentences are further analysable. So, Thesis 2 dictates that their truth-value will be determined by the semantic properties of their constituent expressions. So, we need a theory for the semantic value of expressions other than sentences (including a theory about the truth of quantified sentences).
- Molecular sentences: their truth value is a function of the truth-values of their constituents, given the connectives.

Comment: truth-functionality of logic

Thesis 3: (Extensionality) Substitution of a constituent sentence with another that has the same semantic value will leave the semantic value (i.e., truth-value) of the sentence unchanged.

This is a consequence of Thesis 2. If 3 was violated, then so would 2. (The semantic value of a complex expression would not be fully determined by the semantic values of its parts.)

(Cf. Frege S&R, p.64)

- **Proper Names**

Building a semantic theory for a non-propositional language (atomic sentences have structures)—e.g., subject-predicate form.

What is the contribution of a proper name to the semantic value of the sentences it occurs? N.B. semantic value is truth-value.

"The Thames is a river". If "Thames" was the name of Socrates, the sentence would be false. (Remember the Humpty Dumpty principle.) It's because "Thames" in fact names a river, that the sentence is true.

So, **Thesis 4**: The semantic value of a proper name is the object that it refers to or stands for.

N.B. Frege took the truth-values to be objects. So, sentences are proper names for two objects: True and False. (This is farfetched, however). (cf. "Function and Object", p. 32)

- **Other expressions**

Frege on functions: functions have gaps. Incomplete or unsaturated expressions.

Contrast with proper names and sentences: they have no gaps

- **Predicates**

Numerical predicates: "… is even"
What is the result of slotting a numeral into the gap? A true or a fasle sentence.
Numerical predicates are functions from numbers to truth-values.

Generalise to any predicate, e.g., "… is round".
Functions from objects to truth-values

(Any object can go into the gap of a function.) Frege on objects: "Function and Concept" p.32)

Thesis 5: The semantic value of a predicate is a function.

Intuitively: the semantic value of a predicate is its extension. (The things to which it applies.)

Thesis 6: Functions are extensional: if function f and function g have the same extension, then f=g.

This follows from the thesis of extensionality.

e.g., "4 is even" and "4 is a number divided by two".

- **Logical Connectives**

Functions from truth-values to truth-values

e.g., "-"

(T, F) → (F, T)

"&"

(T, T) → T
(T, F) → F
(F, T) → F
(F, F) → F

- **Quantifiers**

Second order functions

f-o functions: objects as arguments
s-o functions: concepts as arguments

Universal quantifier: ∀x Gx

∀x ( ) is a second-order function from concepts to truth-values
Universe of discourse

"Everyone is mortal"

The concept Gx (… is mortal) is the argument of the function.

∀x Gx
Values: True if G is paired with Truth in the extension of G, that is if every object in the universe of discourse is paired with Truth in the extension of G.

False if at least one object in the universe of discourse is paired with False in the extension of G.

Why second-order? a) ∀x ( ) is true if G is paired with Truth in the extension of G. b) G is paired with Truth in the extension of G if every object in the universe of discourse is paired with Truth in the extension of G.

Existential quantifier: ∃x Hx
"Someone is mortal"

∃x ( ) is a second-order function from concepts to truth-values
∃x Hx
Values: True if G is paired with Truth in the extension of G, that is if at least one object in the universe of discourse is paired with Truth in the extension of G.

False if no object in the universe of discourse is paired with False in the extension of G.

Thesis 7: The semantic value of a predicate is a first-level function from objects to truth-values; the semantic value of a connective is a first-level function from truth-

values to truth-values; the semantic value of a quantifier is a second-level function from concepts to truth-values.

These theses can offer a semantic theory for a language.

Example

Compositional axiom: a sentence coupling a proper name with a predicate is true if and only if the object referred to by the proper name is paired with Truth in the extension of the predicate.
Intuitively: if the object named belongs to the extension of the predicate.
This semantic theory explains how the truth-conditions of a sentence are determined (derived).

Truth-conditions=Structure + reference (semantic value)

Distinguish between truth-value and truth-conditions.
So, here we have the beginnings of a truth-conditional theory of meaning.

Can all sentences be treated by a truth-conditional account of meaning?
Commands, questions
Smart: **compliance-conditions**

Important issue: is just one semantic property (semantic value/reference) enough for a theory of meaning?

❑ Fregean senses ("Sinn")

Reasons for introducing senses.

1. **Empty names**

If the only semantic value of a proper name is the object to which it refers, then empty (e.g., fictional) names are meaningless. But they are not.

So, there must be some other semantic property possessed by a name in addition to having a reference. Frege calls this “sense”.

Cf. The Millian view. (John Stuart Mill)
“proper names are not connotative: they denote the individuals who are called by them; but they do not indicate or imply any attributes as belonging to those individuals”.
The Millian view might be called a "Referential Theory of Meaning" (RTM).
RTM: the meaning of a proper name (or of a singular term in general) is its reference.
But an **RTM** is (seems?) inadequate as a full theory of meaning.

Further problems with RTM (Millian view)

1a. True Existential statements and True Negative Existential statements.
(1) "Bill Clinton exists"
(2) "Pegasus does not exist"

If the meaning of "Bill Clinton" is exhausted by its semantic value (reference), then (1) it tautologically (necessarily) true. For it is meaningful, and hence, on RTM, the name "Bill Clinton" has to have a bearer (referent). Hence, Bill Clinton has to exist. Conversely, (2) would be meaningless. For since "Pegasus" is an empty term, it ends up meaningless on RTM. But (2) is true. Hence--on RTM--we have a paradox: a true sentence can be meaningless.
Here again, proper names can also have sense (even if they have no reference). "Pegasus does not exist" is both meaningful and true, because "Pegasus" has sense. "Bill Clinton exists" is far from being tautological, if "Bill Clinton" has sense as well as reference. For the claim that a name with <u>that</u> sense has a reference is not tautological.

1b. Identity statements
RTM might work in ordinary cases:

(3) "Washington is the capital of USA"

(4) "New York is not the capital of USA"

Since "Washington" and "New York" have different referents, (3) and (4) have different meanings.

But:
(5) "Everest is Everest"
(6) "Everest is Gaurisanker"

(5) is trivial: instance of the law of self-identity.
(6) is informative: two different ways to refer to the same object.

But on RTM, "Everest" and "Gaurisanker" have the same meaning--since they have the same reference. Hence (6) cannot be informative on RTM.
Here again, senses can obviously help show that (6) is informative.

1c. Definite descriptions
What is a definite description? The role of "the".

"The President of USA", "The manager of Manchester United", "My brother", "The shop on the corner", "The author of Das Capital" etc.

Intuitively, DD are like singular terms. (We shall see later that Russell challenged this view.)

Frege's example:

(7) "The morning star is the morning star"
(8) "The morning star is the evening star"

(7) is tautological, whereas (8) is informative (we need to do astronomy to ascertain it). What explains their difference. On RTM, (8) is as uninformative as (7) since "morning star" and "evening star" are co-referential.

Here again, senses help. The meanings of DD are not merely their referential role.

Consider also: (Russell's example)
(9) "The present king of France is bald"

"The present king of France" is a DD. But there is no present king of France (France is a republic). So, on RTM, (9) is meaningless. But (9) is not gibberish. So what does its meaning consist in?

Meinong: possible objects exist as much as actual one.
Frege: this is nonsense. (9) has sense, and hence it's meaningful. (But is it true or false? We'll come back to this.)

2. Opacity of belief contexts (intensional contexts)
Inference (and logic) is extensional (they satisfy Thesis 3).

Everest is the highest mountain.
Everest is Gaurisanker.
Therefore, Gaurisanker is the highest mountain.

This is a valid inference.
But
John believes that Everest is the highest mountain.
Everest is Gaurisanker.
Therefore, John believes that Gaurisanker is the highest mountain.

This is invalid. John might have no views about the identity of Everest and Gaurisanker.

Similarly for Evening star and Morning star.
The example generalises: Thesis 3 (Extensionality) applies to any sentences with the same semantic-value (truth-value).
So, if belief contexts satisfied Thesis 3, then we could say the following:

(10) "John believes that Athens is north of Heraklion"
Since the sentence "Athens is north of Heraklion" has the same truth-value (truth) with the sentence "Jupiter has 8 satellites", we could infer that
(11) "John believes that Jupiter has 8 satellites". But John might have no knowledge of astronomy.

So dilemma: either Thesis 1 has to go or Thesis 3 has to go. So, either the semantic value of a sentence is not its truth-value or extensionality has to be sacrificed.
Frege solves this problem by adding a further semantic property, viz. sense.

3. Informativeness and Understanding

(8) "The morning star is the evening star"

Someone understands this sentence, without knowing its truth-value.

Note: This is unlike (7) "The morning star is the morning star". Someone who understands (7), also knows its truth-value.

If all there was to meaning was semantic value (truth-value) then a) this difference would be impossible and b) (8) would not be really understood, if one didn't know its truth-value. But (8) is understood well. So, meaning cannot just be the semantic value.

Here again, senses helps. One can understand (8) by grasping its sense, without knowing its truth-value.

- **What are <u>Fregean</u> Senses?**

Thesis 8: The sense of an expression is that ingredient of its meaning that determines its semantic value.

Sense determines reference

Thesis 8': If two expressions have the same sense they must have the same reference. Obviously, the converse does not follow. That's the whole point of introducing senses.

A proper name has a semantic value (reference) and a sense, i.e., a property which determines what object it refers to (i.e., what its reference is).

e.g., "Evening star"---> planet Venus

and some descriptive condition which an object has to satisfy in order to be the referent of "Evening star", e.g., "that object that appears in such and such a place in the sky and at such and such times in the evening".

Frege: "A proper name (...) expresses its sense, stands for or designates its reference" (S&R, p.61)

N.B. Here we see the beginnings of descriptive theories of reference (i.e., non-Millian theories).

Thesis 9: It is possible to know the sense of an expression without knowing its semantic value.

Since senses are associated with descriptions, one can understand a description (and hence know the sense) without knowing whether this description is true or false (or, more generally, whether it is satisfied or not).

Thesis 10: The sense of an expression is what someone who understands the expression grasps.

Note: meaning is tied to understanding and grasping.

Thesis 11: The sense of a complex expression is determined by the senses of its constituents.

Compositionality for senses.

Theses 9, 10, and 11 show how senses can solve the problem of informativeness.

It is possible to understand the senses of the expressions "morning star" and "evening star" without knowing their semantic values. Hence (11), it is also possible to understand the sentence "The morning star is the evening star" without knowing its truth-value.

Thesis 12: If someone grasps the senses of two expressions, and the two expressions actually have the same sense, then they must know that the two expressions have the same sense.

This explains why the sentence "The morning star is the morning star" is uninformative. One who grasps the sense of "morning star" knows that "The morning star is the morning star" is trivial and uninformative.
More generally: Thesis 12 suggests that if we grasp senses, then we can grasp facts about synonymy (i.e., facts about identity of senses).

Thesis 13: An expression can have sense even if it lacks a semantic value.

So, empty names can have senses (associated descriptions) even if they have no referent. So, empty names are not necessarily meaningless. (Frege, "On Sense and Reference", p.58).
"In grasping a sense one is not certainly assured of reference".

Thesis 14: A sentence that contains an expression that lacks a semantic value is neither true nor false.

So, "James Bond is the secret agent spy ever" is neither true nor false. This follows from Thesis 2 (compositionality).

Frege's example: "Odysseus was set ashore at Ithaca while sound asleep".

But this is no problem for Frege, since Thesis 13 guarantees that this sentence can have sense (and hence be meaningful). (In fact, it satisfies Thesis 11--compositionality of senses--so it has to have a sense.)

So, unlike Meinong, Frege does not need to introduce possible objects. The expression “The present king of France is bald” is meaningful, yet it is neither true nor false.

How can senses solve the problem of opacity?
Do we have to discard Thesis 3 (substitutivity)?

(12) "John believes that Bob Dylan is Bob Dylan".
(13) "John believes that Bob Dylan is the Robert Zimmerman".

Frege's solution: within opaque contexts (intensional contexts) names refer to their senses (their customary senses). So, they are not co-referential.

So, in such context, "Bob Dylan" refers not to the man but to the sense of the name "Bob Dylan". So, in moving from (12) to (13) we haven't substituted one co-referential expression for another. So, the principle of substitutivity is not violated.
More generally:
John believes that Everest is the highest mountain.
Everest is Gaurisanker.
Therefore, John believes that Gaurisanker is the highest mountain.

This argument is invalid because the terms "Everest" and "Gaurisanker" have different semantic values, on the one hand, in premise two and, on the other hand, in premise one and the conclusion of the argument.

Moral: an expression does not have the same semantic value (reference) in all contexts. When it occurs in belief-contexts, its sematnic value is not its not its (usual) reference but its' sense.
This happens in natural language. In a perfect logical language ambiguity has to be excised.

Thesis 15: In a belief context, the (indirect) referent of a proper name is its customary sense.

See Frege Sense and Reference pp.58-9 and 65.

Thesis 16: Substitution of one expression in a sentence by another which has the same sense will leave the sense of the sentence unchanged.

This follows from Thesis 11 (compositionality of senses).
cf. Frege S & R, p.67

So, the substitutivity principle is not violated. In fact, it is explained why the sentences
(10) "John believes that Athens is north of Heraklion"
and
(11) "John believes that Jupiter has 8 satellites" are not equivalent, although "Athens is north of Heraklion" and "Jupiter has 8 satellites" have the same truth-value (truth). They cannot be interchanged in belief contexts because they do not have the same sense. (They don't express the same thought).
But take: "Heraklion is south of Athens". Now, arguably, this has the same sense as "Athens is north of Heraklion". Hence, it can be substituted for the latter in (10).

Intuitively, "Heraklion is south of Athens" and "Athens is north of Heraklion" express the same thought.
Conversely (Frege): "The morning star is the body illuminated by the sun" expresses a different thought from the sentence "The evening star is the body illuminated by the sun", although they have exactly the same semantic value. The express different thoughts, therefore they have different senses.

Thesis 17: **The sense of a sentence is a thought**.

N.B. Frege was an anti-psychologist. Thoughts (and hence senses) are objective--they specify a condition in the world such that when it obtains it is necessary and sufficient for the truth of the sentence that expresses the thought.

Thoughts are not psychological or subjective.

Thesis 17': **The sense of sentence is its truth-condition.**

N.B. Thoughts, of course, are not truth-conditions. But thoughts have truth-conditions and, in this sense, the sense of a sentence is the truth-condition of the thought it expresses.

Comment:
We said before (Thesis 8):
Sense is this ingredient of meaning which determines its semantic value. So, sense determines reference.
Thesis 10 says that the sense of an expression is what someone who understand this expression grasps.
If we put them together we get: the sense of an expression is this ingredient of its meaning which someone who understands this expression grasps and thereby determines the expression's semantic value.
This might seem to create a tension: does someone who understands an expression has to know its semantic value (i.e., whether it is true or false)? If so, we would have a contradiction in our theory (Thesis 9).

Thesis 18: The semantic value of an expression is no part of what someone who understands the expression grasps.

This thesis ensures that there is no tension in the theory.
Besides: it guarantees the directedness of thought to the world.

Frege: "A truth value cannot be a part of the thought, any more than, say, the Sun can, for it is not a sense but an object" (S&R, p.64).

"the mere thought alone yields no knowledge, but only the thought with its reference, i.e., its truth value. Judgements can be regarded as advances from a thought to a truth value" (S&R, p.65).

Corollary: insofar as meaning is tied to grasping something (a sense), the semantic value of an expression is not an ingredient of meaning.

Frege: "only the sense, not the reference, of the part is relevant to the sense of the whole sentence. The thought remains the same whether 'Odysseus' has reference or not" (S&R, p.63).

Yet, meaning issues in reference. The point of the corollary is this:
Dummett: "the understanding which a speaker of a language has of a word in that language ... can never consist merely in his associating a certain thing with it as its referent; there must be some particular means by which this association is effected, the knowledge of which constitutes his grasp of its sense".

Frege: "It is the striving for truth that drives us always to advance from the sense to the reference". (S&R, p.63)

- **More on what senses are.**

Frege never defines 'sense'

1. Frege's critique of Locke: Senses are not subjective ideas.

Lockean view: association of an expression with an idea (a mental picture, image).
Essay pp.207-8

public nature of language vs. private nature of ideas.
different speakers can assign the same meaning to their words; they know what other people mean by their words; communication requires common senses.

But ideas are private. How can communication be possible on this view?
Locke p. 208 (III.II.2)
How can we know that your private idea of "blue" is the same as mine?

Locke on the possibility of "inverted spectrum". (Essay, pp.195-196)

Frege on "Bucephalus". (Sense and Reference, p.59)

A sense may be "the common property of many and therefore is not a part or a mode of the individual mind" (S&R, p.59).

same sense--->different ideas

Senses can be shared by speakers of the language whereas ideas cannot.

Locke on "secret reference". (We suppose that our ideas are the same. But this is irrelevant for even if they were not, we would manage to communicate.) Essay p.209 (III.II4-6)

Thesis 19: Sense is objective: grasping a sense is not a matter of having ideas, mental images, or private psychological items.

Comment: Can't a word have different senses for different speakers of the language? (e.g., 'Aristotle'). This can be tolerated in natural language but not in logically perfect languages.

2. Sense is normative

Thesis 20: The sense of an expression is normative: it constitutes a normative constraint that determines which uses of this expressions are correct and incorrect.

3. Senses as "modes of presentation of what is designated".

cf.

The teacher of Alexander the Great

The founder of the Lyceum

or

The husband of Penelope

The father of Telemachus

"Mode of presentation" suggests that a name has a referent which is presented in different ways (senses).

- **Problems with Sense**

1. In what sense are senses objective?

N.B. same referent, different modes of presentation. e.g., "Aristotle"

The teacher of Alexander the Great
The founder of the Lyceum

If speakers associate different descriptions with a name, i.e., if they express different thoughts (their sentences have different senses) in what sense can we say that they manage to communicate? Frege seems unable to establish that they communicate--which seems absurd. They do!

Don't Lockean problems arise for Frege too?

Frege (S&R, p.58): this can be tolerated. Yet, it makes possible that we do what we currently do with our language without really communicating with each other.

2. Kripke's objection (to be presented later on).

3a. Problems with Thesis 13. Can there be sense without the thought having a truth-value?

- Empty names.

Frege's example: "Odysseus was set ashore at Ithaca while sound asleep". This expresses a thought, but it lacks truth-value.

For some this is absurd. A determinate belief must have a truth-value.

3b. Problems with Thesis 13. Sense as mode of presentation. If there is no referent (e.g. empty names), how can there be a mode of presentation for it?

The husband of Penelope
The father of Telemachus

Suggestion: sense as a mode of determination of reference.
Sense are routes to reference, and it may well be the case that there is nothing at the end of the journey.

mode of presentation= mode of determination + something else (the object being appropriately present).

4. Does Frege offer a correct theory for the sense of simple proper names?

If mode of presentation is a kind of definite description, then he might have to make one of the 'privileged': the one that presents the sense of the name. But how is this to be done?
So, is the alternative to take the reference of the simple proper names as their only semantic value?

III. DESCRIPTIVE THEORIES OF MEANING (RUSSELL)

- Russell's two main works on meaning:

1. "On Denoting" (1905)
2. "The Philosophy of Logical Atomism" (1918).

Early Russell: The meaning of a proper name is to be identified with the object that the name denotes.
Proper names: (by definition) words for particulars

What are particulars? Terms for relations in atomic facts (PLA, p.199).

Fa, Rab, Gabc etc. "a", "b" etc. are names; "F", "R", "G" etc. are predicates.

Russell's ontology circa 1918: Wittgensteinian--as in the Tractatus. Facts: Facts are what make statements true or false.

"We express a fact, for example, when we say that a certain thing has a certain property, or that it has a certain relation to another thing" (PLA, p. 183).

atomic, molecular
particular facts, general facts

Proper names do not describe particulars--they name them. (This follows from the thesis that names have only denotation, not connotation. This is a Millian view.) But, as we shall see, Russell is not a Millian about proper names, since he denies that there are genuine proper names.
Particulars: they are completely self-subsistent. ("Each particular that there is in the world does not in any way logically depend upon any other particular" PLA, p. 202.)

Russell's epistemology: "in order to understand a name for a particular, the only thing necessary is to be acquainted with that particular".

knowledge by acquaintance vs knowledge by description.

(things we have presentations of vs things we only reach by means of denoting phrases--OD, p. 41)

Naming is tied up to acquaintance.

"A name--in the narrow logical sense of a word whose meaning is a particular--can only be applied to a particular with which the speaker is acquainted, because you cannot name anything you are not acquainted with" (PLA, 201).

So, understanding a name → acquaintance with a particular (that which is named).
This is a "full, concrete and adequate" understanding of a name.

Two issues that crop up:

1. Are there really any proper names (in the logical and not everyday sense of the word)?
2. Are definite descriptions (which seem to have the surface grammar of a name) really names? (Are they singular terms?)

Let's start with the second question first.
Denoting phrases such as "the author of Waverley", or "the President of the USA" etc.
Let's call them Definite Descriptions (DDs).

Russell: they are not names--taken in isolation, they don't name anything in particular.

a. A name is a simple symbol; DDs are complex expressions
b. A name is understood independently of other words; DDs are understood by understanding the meanings of their constituents.

"To know the meaning of a name is to know who it is applied to" (PLA, p.244).

c. (A): "Scott is the author of Waverley"

Doesn't (A) assert that "Scott" and "the author of Waverley" are two names of the same object?
Russell: this is a delusion

N.B. "is" might be short for "is identical to" (so the "is" of identity)
or it might be the "is" of predication (as in "Socrates is mortal")

In (A) the "is" is the "is" of identity. So, if "the author of Waverley" refers to Scott, and if it is a proper name, then its denotation is Scott. Then (A) should be equivalent in meaning to (B): "Scott is Scott"--which is a tautology. But "Scott is the author of Waverley" is not a tautology. So, (A) and (B) are not equivalent in meaning. So, "the author of Waverley" cannot be a proper name.

Russell: when you substitute a name for a description, the result is always a different proposition. (PLA, p.247)

Repeat: If DDs are treated as names, then their meaning should be the object which they denote. Then what is meant by saying that Scott is the author of Waverley should be identical with what is meant by saying that Scott is Scott. But this cannot be so. For when King George IV wished to know whether Scott was the author of Waverley he surely didn't wish to know that Scott was Scott.

d. "Plato is Plato" (this is a tautology--an instance of the law of self-identity--x=x). But what about "The author of Waverley is the author of Waverley"? Is it a tautology? Not necessarily, since "the author of Waverley" is not a name. If it was a name, then we could substitute "Scott" for "the author of Waverley". But "Scott is the author of Waverley" and "The author of Waverley is the author of Waverley" are different propositions.

Another example:
"The present King of France is bald"

The law of excluded middle holds: The present King of France is either bald or not bald. Yet, since there is no present King of France to be either bald or not bald, we face a dilemma: either expressions such as "The present King of France" violate the law of excluded middle, or they are not denoting expressions (i.e., they are not names).

Russell opted for the latter horn.

Definite (as well as indefinite) descriptions are not names.

Incomplete expressions (symbols): no meaning in isolation--unlike proper names which have a meaning in isolation.

Incompleteness: 1. they do not need to have denotation (reference/semantic value)
2. the meaning of any sentence in which an incomplete symbol occurs can be analysed in such a way that the resulting sentence no longer contains the symbol or any synonym for it.

How then should we analyse them? (Get their proper logical form, not their superficial grammatical form).

Indefinite descriptions. e.g., "A man walked in the room"

There is a x, (man x) and x walked in the room.
The indefinite description disappears. All that remains is a general existential statement.

Definite descriptions: add uniqueness.
"Scott is *the* author of Waverley"

There is at least one thing that authored Waverley.

There is at most one thing that authored Waverley.
This thing is identical with Scott.

There is an x (authored Waverley, x) and for all y (if authored Waverley, y) then x=y and (Scott=x).

N.B. This expression does not contain any constituent 'the author of Waverley' for which we could substitute 'Scott'. The so-called denoting expression disappears.

Quine: "the burden of reference which had been put upon descriptive phrases" is taken over by quantifiers". (On What There Is, p.194).

So, "The present King of France is bald" is just false--for Russell.

There is an x (kings France now x) and for all y (if kings France now y) then x=y and (bald x).

Since, nothing satisfies this sentence, it is false. (Remember, Frege thought that it lacks truth-value.)

The use of a description carries the covert assertion that there is an object that satisfies it. But, according to Russell's method, the description is broken up and disappears. No constituent of the expression that takes the place of a description corresponds to the descriptive phrase.

What if a DD stands as a subject in a sentence?

"The author of Waverley is human" (where, "is" is the "is" of predication).
DD is not the subject in the sense in which a proper name is.

There is at least one thing that authored Waverley.
There is at most one thing that authored Waverley.
This thing is human.

There is an x (authored Waverley x) and for all y (if authored Waverley y) then x=y and (human x).

Clearly, this sentence doesn't contain the expression "The author of Waverley". So, "the author of Waverley" is not its subject. Besides, "The author of Waverley is human" makes no reference to Scott.
Does the description "The author of Waverley" stand for Scott? No. If it did, we could replace it by "Scott". But as we have seen, "Scott is the author of Waverley" is informative, whereas "Scott is Scott" is a tautology.
But, Russell adds, the expression "The author of Waverley" doesn't stand for anything else either. (If it did stand for something else, the sentence "Scott is the author of Waverley" would be false--yet, it is true.
So Russell concludes that the expression "The author of Waverley", taken in isolation, stands for nothing. (Incomplete symbol).
But within a sentence--i.e., in context--it contributes to the meaning of the sentence, if properly analysed: for the sentence is either true or false.

Back to the first question above:
Are there really any proper names (in the logical and not everyday sense of the word)?

If there were any genuine proper names, they would be different from descriptions. Only genuine proper names (i.e., those that have no connotation) cannot be treated as descriptions--and hence cannot be analysed away by Russell's method.
So, take "Scott" to be a genuine proper name. And take "Sir Walter" also to be a proper name. Then, when we say that "Scott is Sir Walter", we really name the same object a. So, all we are saying is that a is self-identical. But surely, we mean to say much more when we say that "Scott is Sir Walter". For this is an informative statement.
So, the only way to show that statements such as "Scott is Sir Walter" are informative, is to treat "Scott" and "Sir Walter" not as genuine names, but as abbreviated descriptions.

"The person called 'Scott' is the same as the person called 'Sir Walter'"

The expression "The person called 'Scott'" is an abbreviated description. It has connotation.

So Russell concludes that all names are really abbreviated descriptions. "The manes that we commonly use, like 'Socrates', are really abbreviations for descriptions" (PLA, 200).
So: the person called 'Socrates' is whoever satisfies a certain description. (The description is the connotation of the term, and since the term has connotation, it cannot be a proper name.)
Repeat: For Russell, ordinary proper names ("Plato", "London") have connotation--so he treats them as implicit descriptions.

Ordinary proper names are abbreviated descriptions.

Genuine proper names have no connotation: they simply stand for an object.
Their use guarantees that the objects they stand for exist.
Example:

"Romulus existed"

perfectly significant statement. But what does it mean?

(A): "x has such-and-such properties" (those attributed to Romulus)

That Romulus exists means that there is a (unique) person that satisfies this description (i.e., that makes (A) true).
That Romulus didn't exist means that there is no value of x that makes (A) true.

So Russell concludes that "Romulus" is not a name but a "truncated description". It stands for whatever did so-and-so.
So, the only genuine proper names are what Russell calls "logically proper names": "this" and "that". They refer to objects of our direct acquaintance: feelings, sense-data.

So: Russell and Frege disagree on proper names. Russell: if they are really proper names, then they have no sense--Frege: they do. But Russell thinks that the only really proper names are the logical proper names.

Far-reaching implication of Russell's views:

- **meaningful talk does not presuppose existence**
- **a singular term need not name to be significant.**
- **existential commitments are through variables not through names.**

Problems with Definite Descriptions

- **Strawson "On Referring" (Mind, Mind, 59, 1950).**

Russell's insistence on logical form loses contact with the use of language.

1. Expressions do not refer. Speakers use them to refer.
Strawson denies that expressions have semantic values independently of the uses to which speaker's put them and the contexts in which they occur. So, Strawson favours a theory of Speaker Reference.

Semantic reference: meanings are invariable relations between words and the world.
Speaker reference: meanings depend on (vary with) the context of use. Reference is not an abstract relation between a word and the world, but an act.
Meaning attaches to sentences, but truth and falsity attaches to **utterances**.

"The present King of France is bald"

How would we react if we heard this sentence?

Russell: analyse it's logical form and declare it meaningful but false.
Strawson: Depends on the context of its utterance. It's meaningful, because it could be true or false. But if it is uttered now, it's not just false. It's inappropriate (it has

misfired). If it was uttered in 1788, it would be false. And if Luis XIV was indeed bald, it would be true.

Moral: singular noun-terms (and descriptions) can be used to refer to different things on different occasions (contexts). A meaningful sentence can be uttered in a context to make a true assertion. Or it may be uttered in a different context to make a false assertion. Or it may be uttered in a context to make an assertion that it in neither true nor false.

Quine's distinction is relevant here: standing sentences vs occasion sentences

"The milk is white"
All of its utterances are either true or false independently of the context.
"It's raining".

The truth or falsity of an utterance depends on the context.

1a. "The table is full of books"

Russell's analysis of DDs seems to yield the wrong results. Does it not entail the uniqueness of the table? (There is exactly one table ...). Here again context disambiguates.

- **Donnellan "Reference and Definite Descriptions", Philosophical Review, 75, 1966.**

Russell failed to distinguish between attributive uses and referential uses of DDs.

"Smith's murderer is insane"

Attributive uses (like Russell): whoever committed this crime is insane.

But suppose that a person is in the dock being charged with Smith's murder. And suppose that I am convinced that this is the person that committed the crime. Then I use "Smith's murderer is insane" referentially to refer to this particular person.

Russell overlooked the referential use of DDs
But Strawson overlooked the attributive use. He speaks as if all uses of DDs are referential, used in a context, to draw the attention to a particular person, place or thing.

How to distinguish the two uses?

"A speaker who uses a definite description attributively in an assertion states something about whoever or whatever is so-and-so." (Donnellan, p. 46)

In this case the definite description occurs "essentially"; the speaker wishes to assert something about whoever fits the description.

"A speaker who uses a definite description referentially in an assertion [...] uses the description to enable his audience to pick out whom or what he is talking about and states something about that person or thing"

The referential use is merely "a tool for doing one job--calling attention to a person or thing--and in general any other device for doing the same job, another description or a name, would do as well".
In the attributive use "the attribute of being so-and-so is all important, while it is not in the referential use".
There is another important difference. Suppose that nothing satisfies the description (e.g., suppose that Smith committed suicide).

"The o is x."

Then, on the attributive use, nothing has been said to be x. But on the referential use, even if nothing is o, it's not the case that nothing has been said to be x.

"Her husband is nice to her".
Even if she is not married, on the referential use of "her husband" it has been said of the person who is nice to her that he is nice to her.

Interesting case:
"Who is the interesting man drinking martini?"

Suppose that the man we have in mind does not drink martini but vodka, and that another man (Dino) nearly drinks martini. Whom are we referring to? The interesting thing is that we might refer to the man we have in mind (and not to Dino), by asserting something false of him.
"near-miss" cases. In such cases we succeed in referring because our hearers understand our intentions even though what we are uttering is actually false.

- **Problems with Russell's theory of names.**

1. Is the meaning of a proper name to be identified with the object the name denotes?

The name seems to make a contribution to the meaning of the sentences it occurs irrespective of whether it denotes something or not--this is, in essence, Frege's objection.
Of course, Russell denies that what we take to be proper names are genuinely proper names. And since they are abbreviated descriptions, one can think of the descriptions as expressing the sense of a name.

2. Unfortunate consequence: the meaning of logically proper names ("this", "that") differs almost on any occasion in which they are uttered.

3. Can we do without senses?

"Common words, even proper names, are usually really descriptions. That is to say, the thought in the mind of a person using a proper name correctly can generally only

be expressed explicitly if we replace the proper name by a description. Moreover, the description required to express the thought will vary fro different people, or for the same person at different times. The only thing constant (so long as the name is rigidly used) is the object to which the name applies. But so long as this remains constant, the particular description involved usually makes no difference to the truth or falsehood of the proposition in which the name appears" (Problems of Philosophy, 29-30).

Russell's example: "Bismarck"
"the first Chancellor of the German Empire"
"the most powerful man in Europe"
"Bismarck was the most astute diplomat"

If these different DDs are associated with "Bismarck", how can we manage to communicate?
Russell: When I try to communicate, I don't try to share my thoughts with you, but rather to make you have a thought about Bismarck (however you think of him) that he was the most astute diplomat. So for Russell it is enough that the various descriptions stand for the same thing.
But can this be right? Suppose that I associate with "Bismarck" the description "player of the German national team". This is wrong, but if I tried to communicate with you (who associate with "Bismarck" a correct description) wouldn't I succeed in conveying to you a thought about Bismarck? Note: the two descriptions do not stand for the same thing. So, Russell's account fails. We manage to communicate even if the various descriptions fail to stand for the same thing.
This seems to take us back to Fregean senses. What matters for communication is not the object that descriptions stand for but the senses associated with our expressions. (This does not mean that we have solved the problems that Frege faces with senses.)

4. Which description should we privilege and why?

5. Genuine disagreement becomes equivocation.

Me: "Quine is the author of 'Word and Object'"

You: "Quine is the most famous Harvard philosopher"

Suppose I say "Quine was American" and you deny this? We disagree. But not so if we treat names as abbreviated descriptions à la Russell.

"There is one and only one x such that x authored 'Word and Object' and x was American"
is perfectly consistent with
"There is one and only one x such that x was the most famous Harvard philosopher and x was not American".

- **Searle to the rescue: Cluster theory**

No need to associate a name with a particular description. Instead a name refers to whatever objects satisfy a sufficient but vague and unspecified number of the descriptions generally associated with it.

Vagueness is crucial.

- **Kripke's critique of description theories**

1. Description theories render analytic statements that ought to be contingent. If the meaning of "Aristotle" is given by the description "The pupil of Plato and the teacher of Alexander the Great" then,

"Aristotle was the pupil of Plato and the teacher of Alexander the Great"

is analytically (necessarily) true.

But it is possible that Aristotle never met Plato, Alexander etc.

2. Central thought of Descriptive theories: the competent speaker must know some identifying facts about the referent in order to refer successfully to it. These identifying facts are captured by descriptions.

- **Descriptions are neither necessary nor sufficient for fixing the reference of a name.**

Kripke: it is neither necessary nor sufficient for a name to designate an object that the speaker associates with the name identifying knowledge of the object.

- Description are not sufficient for reference (Something might satisfy the description without being the reference)

Goedel/Schmidt case.

If Schmidt had proved the Incompleteness theory, when we used the name "Goedel" would we refer to Goedel or to Schmidt?

"Einstein invented the atomic bomb". Does this sentence refer to Oppenheimer?

Even where descriptions that identify an object are associated by a speaker with a name, the name may not refer to that individual.

- Description are not necessary for reference (Something might be the referent without satisfying the description)

"Einstein" refers to Einstein even if people associate wrong descriptions with the name or even if they are ignorant about the relevant descriptions.

What if all of out relevant beliefs are false? We can still refer to a person. e.g., "Moses".

We can use a name to designate an object when we are ignorant of the object. We can do it even when our error about the object is massive.

IV. CAUSAL THEORIES OF REFERENCE (KRIPKE-PUTNAM)

❖ MOTIVATIONS AGAINST DESCRIPTION THEORIES

i. Proper names as **rigid designators**: the designate the same individual in every possible world.
They are different from DDs in exactly this respect.
DDs do not name the same individuals in all possible worlds.
e.g., "The President of the USA"

Consider: "The President of the USA might not have been the President of the USA"
Perfectly sensible.
But compare: "George W. Bush might not have been George W. Bush"
It hardly makes sense.
Of course, this is to be distinguished from the correct claim that George W. Bush might not have been named "George W. Bush".

ii. A Description theory of reference does not explain how language hooks on the world.
Some terms cannot have their reference fixed by description.
Descriptions link words with other words.

iii. Semantic externalism
The reference is part of the meaning of a word.
Putnam's Twin-Earth story.

Meaning cannot just be association of words with words--associations can be the same and yet the reference can be different

e.g., 'water' on Earth and 'water' in Twin-Earth

Meanings ain't in the head: no internal state of the speaker that determines the reference

- **CAUSAL THEORY I: PROPER NAMES**

The Causal Theory of Names--introduced by Kripke--identifies the semantic value (denotation/reference) of a name with the individual/entity that this names refers to. In order to get a proper philosophical theory of reference we need to deal with two issues:

- **How reference is fixed (i.e., what makes it the case that the name stands for this rather than that individual)**

e.g., how is it that "Aristotle" stands for Aristotle and not for Plato?

- **How reference is transmitted through a linguistic community (what makes it the case that, e.g., by using the name "Aristotle" I refer to Aristotle and not to Plato).**

Reference-fixing

What fixes the reference is a causal--historical chain which links the current use of a name with an introducing event; that is, an event which associated the name with its referent. This introducing event Kripke called "act of baptism", where a certain name was picked to dub an individual. Descriptions associated with the name might (all) be false, and yet the users of the name still refer to the individual dubbed, insofar as the their use of the name is part of a causal transmission-chain which goes back to the dubbing ceremony.
Occasionally, the introducing event may involve some description of the individual introduced. In fact, there are cases in which the introduction of an entity, or an individual, is made only via a description, e.g., the introduction of the name 'Neptune' for the planet Neptune, or of the 'Jack the Ripper' (cf. Kripke, 1972: 79-80 & 96). But the description is **not** analytically tied to the term, as traditional description theories would have it. The description just "fixes the reference by some contingent marks of

the object" (Kripke, op.cit.: 106). As a rule, however, the causal theory insists that 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 in the dubbing ceremony.

Reference-transmission

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.
So, the thrust of the theory is that the relation between a word and an object is **direct**--a direct causal link--unmediated by a concept. In particular, the causal theory dispenses with senses as reference-fixing devices and suggests that the reference of a word is the entity which "grounded" the word in the dubbing ceremony in which the word was first introduced. The superiority of causal theories when it comes to proper names is well taken even by some of the most incisive critics of the theory (cf. Unger 1983).

- **CAUSAL THEORY II: OTHER TERMS**

Generalisation of Kripke's theory.

Natural-kind terms

As Putnam observed, Kripke's (1972) theory of proper names offers a model for a theory of reference of 'natural-kind' terms which feature in folk and scientific theories: the reference of a natural-kind term (e.g., "water", or "tiger") is fixed during an introducing event, i.e., an event during which the term is attached to a substance, or a kind, where samples of this substance or instances of this kind are present and ground the word. According to Putnam, reference is fixed by "things which are given existentially" (1983b: 73). In the case of a natural-kind term, e.g., 'water', or 'tiger', this means that one picks out by ostension an object, attaches a name to it, and then asserts that this name applies to all and only the objects that have **the same nature** as the one present in the introductory event.

A guide to the sameness of nature is the similarity that there is among the sample (or instance) which is present and other items of the same kind. So,

for all x, x is an elephant iff x stands in a specific similarity relation to this specific object, picked out by ostension

and for all t, t refers to elephants (i.e., the extension of t is the class of elephants) iff t refers to this specific object (this elephant), picked out by ostension.

As Putnam put it, "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.

Physical magnitude terms

Fixing the reference of 'physical magnitude' terms is a variation of the same theme: when confronted with some observable phenomena, it is reasonable to assume that there is a physical magnitude, or entity, which causes them. Then we (or indeed, the first person to notice them) dub this magnitude with a term t and associate this magnitude with the production of these phenomena. This is the introducing event of the term t as referring to this magnitude. One will, typically, surround the term with a description, i.e., with a causal story, of the nature of the posited magnitude and of the properties in virtue of which the magnitude causes its paradigmatic observable effects. This initial description will, most likely, be incomplete, or even misguided. It may even be a wrong description, a totally mistaken account of the nature of this causal agent. But, on the pure causal account, one has nonetheless introduced existentially a referent--an entity causally responsible for certain effects to which the term t refers. That is, one has asserted that:

There is an α {α is causally responsible for certain effects F and for all t (t refers to α if and only if x picks out the causal agent of F)}.

Then, the thrust of the causal theory of reference is this: it is one thing to fix the reference of a term **t**; it is quite another matter to find out the exact nature of the referent α of **t** and hence the correct description associated with **t**. Our beliefs about this entity may change, according 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).

Success Stories

1. Senses

2. No analytic link between descriptions and reference.

2a. Different speakers can use different descriptions and yet refer to the same individuals.

2b. No need to identify the bearer of the name in order to refer to it.

2c. Ignorance and error are no bars to reference.

Yet, **2d**: the causal theory can explain the use of descriptions to identify a referent--without of course making the name synonymous with a description.

3. Putnam on a four-dimensional meaning theory: Stereotype, semantic marker, syntactic marker, and "the dots of extension".

4. How does the theory deal with the traditional problems?

a. Identity statements

Everest is Everest

Everest is Gaurisanker

The terms "Everest" and "Gaurishanker" refer to the same object but they are grounded by different causal networks. So, second statement is informative, since a

speaker should have access to the different causal networks in order to affirm it as true.

b. Intentional contexts

Tom believes that Bob Dylan was the greatest singer ever.
Tom believes that Robert Zimmerman was the greatest singer ever.

The two beliefs are grounded in different causal networks.

c. Empty names
"Odysseus"

Empty names are also grounded in causal networks, even though the causal network is not properly grounded in an object.

d. Existence Statements

"Odysseus does not exist".

Both meaningful and true. Meaningful because the name is grounded in a causal network and true because the network is not grounded in an object.

5. Solving the alleged incommensurability.

- **Problems for the Causal theory**
  - **Change of reference**

"Madagascar".

  - **Problems with similarity**

According to the causal theories, natural-kind terms get their reference during an introductory event in which some instances of the kind (or a sample of it) are present. To a first approximation, it is correct to say that ostension allows us, in principle, to

introduce names for kinds prior to having any knowledge of what makes them into a kind. We do that by appealing to situations in which we believe that such kinds are exemplified (cf. Boyd, 1993: 492). But ostension is not enough to fix the reference of a kind-word unambiguously. It will often establish referential connections between a word and more than one kind. At any rate, there is no guarantee that it will not. Besides, ostension brings us in contact with a sample, or an instance, of kind, and not with its extension as a whole. But the word does not refer only to the sample present. It refers to everything that belongs to its extension. What exactly, then, 'binds together' the sample, or the instances, present at the introductory event, and the other items that belong to the extension of the term? Ostension cannot possibly do the trick. So, reference-fixing should involve more than ostension. Devitt and Sterelny (1987: 72) suggest that it should also involve a "structural component": the extension of a kind-term includes all and only items which have the same internal structure as the ostensively given samples. Manifest properties are not robust enough to circumscribe the boundaries of the kind. Causally relevant differences as well as similarities in the behaviour of two items that are grouped together into an "intuitive kind" (cf. Quine, 1969: 40) are typically due to some internal structural differences or similarities. Ice is a kind of water not because of its manifest properties but because of its internal structure. And a liquid which has the appearance of water might well kill you if you drink it, unless it is H_2O.

Once, however, an appeal to the internal structure is made, what bears the burden of reference is some parts of the theoretical environment in which the term is embedded. An item is said to belong to the extension of a kind-term because it is relevantly similar to the samples present in the introduction of the term. But the properties in virtue of which it is deemed to be relevantly similar to the known samples are those specified in the theoretical description of the internal structure of the samples. In fact, it is these properties which determine that there is a natural kind to which these samples, as well as the item at hand, belong. What, if anything, binds together all samples or items we have encountered, as well as others we have not, into a natural kind is the fact that they all have the same internal structure. But there is no theory-independent way to specify what constitutes the internal structure of the members of a kind. It is theoretical descriptions of the internal structure which show that the extension of the kind-word is not gerrymandered. If we did not rely on such theoretical descriptions of the internal structure to fix the reference of a kind-word, then we would have no way to argue that

the extension of the kind-word is a natural kind. An example will help to drive the point home. Causal theories are certainly right in saying that 'water' refers to the substance with chemical structure H_2O. But this assertion amounts to admitting that the manifest properties of the samples we encounter are not sufficiently robust to determine the extension of the kind-term 'water'. Instead, the unobservable properties of the samples should be appealed to. But to claim that the unobservable structure of the samples is H_2O is to adopt a theory about the structure of this substance. Hence, what bears the burden of reference of 'water' is part of the theoretical environment in which the word is embedded.

Causal theorists might be content with rehearsing their point that, still, what determines reference is not beliefs or theoretical descriptions, but the internal structure, whatever that is: our beliefs about, and descriptions of, the internal structure may all be wrong and yet 'water' still refers to whatever substance exhibits certain manifest properties of the samples used to introduce the term 'water'. Even so, some theoretical descriptions (or beliefs) are necessary, for instance that there is a unique substance (a natural kind) which exhibits all these manifest properties, and that it does so because of its (unknown) internal structure. But the issue is more complicated than that. Suppose, for instance, that we discovered that the chemical composition of the samples used to introduce the term 'water' is not H_2O. (If you think that this is a science-fiction story, then you may think of phlogiston instead of water.) Then, there are two options available to the causal theory: either that the term 'water' does not refer to water, or that it still does. The move favoured by pure causal theories would surely be to say that 'water' still refers to whatever it is that has the manifest properties of the samples which were present at the introduction of the term 'water', even though the real constitution of the samples is not H_2O. This move, however, makes it vivid that the pure causal theory should get its priorities right. For, the causal theory had to put the internal structure of the posited kind ahead of the manifest properties of some samples in order to argue that the similarity among these samples is robust enough to warrant positing a natural kind--water-- to which the term 'water' refers. But when faced with a misidentification of the internal structure of the posited kind, the causal theory has to reverse the order and put the manifest properties ahead of the internal structure in order to argue that samples (or items) which share these properties are nonetheless sufficiently similar to be grouped together as a natural kind with some

internal structure or other, even though they do not have the internal structure we thought they did. These reversal of priorities seems, to say the least, ad hoc. For, if the internal structure, as specified by our best theories, is appealed to to posit and identify the referent of a natural-kind term, then it should also be put ahead of the manifest properties when it comes to a misidentification of the referent of a natural-kind term. To think otherwise is to think that manifest properties are robust enough to determine natural kinds, a thought which we saw is problematic. I think the problem just discussed leaves the causal theory with a pressing dilemma. If the internal properties are consistently put ahead of manifest properties in positing a kind as natural and in taking it to be the referent of a natural-kind term, then when faced with a misidentification of the internal structure of the referent, the causal theory has either to perform a reversal of the roles of internal and of manifest properties in reference-fixing, or else it has to concede that the natural-kind term whose referent was misidentified by our current best theoretical descriptions does not refer to anything. Opting for the first horn of the dilemma would be unmotivated and ad hoc, while opting for the second horn would amount to admitting that theoretical descriptions do play a central role in reference-fixing. The upshot, I think, is that the causal theory should take the role of theoretical descriptions of the internal structure of a posited kind seriously. Positing a natural kind with a certain internal structure should not be a mere place-holder for whatever theoretical description of this "existentially given" kind will be spewed out by ideal science. Rather, positing a natural kind with a certain internal structure should be tied to a description of its properties--a description, that is, of what this internal structure is--in such a way that if there is no kind which has these properties, then we may have to just admit that a word which was taken to refer to this kind does not, after all, refer.

- **The qua problem**

The very idea of a paradigm case of a member of a natural kind entails that we have, at least some knowledge of what this natural kind is, and in particular of in virtue of what this specimen is a paradigm case of a kind, as well as some knowledge of the fact that this is a natural kind -and not an artefact- after all. For instance, when we say that the elephants kept at the London Zoo are paradigm cases of elephants, we classify them qua members o f a natural kind. We therefore associate with them a natural kind

term and not, say, a political term. We classify them qua elephants and not qua mammals, at least primarily. We also attach an animal term to them and not, say, a mineral term. We examine other elephants qua similar to the paradigm elephants. These consideration are known in the literature as 'the qua-problem'. (cf. Papineau, ([1979], ch.5, sect.5, esp. pp.158-162 and Sterelny & Devitt, [1987], pp.72-75) They show that successful reference-fixing cannot be so belief- blind as it was thought initially. That is to say, there are some indispensable background (correct) beliefs we seem to possess and (need to) use in reference-fixing.

- **Can there be referential failure?**

As noted earlier, the causal theory suggests that reference is fixed by things which are given purely existentially. We also agreed that when it comes to positing of observable natural kinds, ostension does play some role in fixing the reference of natural-kind terms. But when it comes to the reference of theoretical terms, ostension cannot be of any help at all. When, for instance, Benjamin Franklin introduced the term 'electricity' what he offered was something like this: the phenomena of sparks and lightning bolts indicate, or suggest, that there is a physical magnitude which causes them--adding, possibly that their cause, what he called electricity, is a substance capable of flow or motion (cf. Putnam, 1975a:199). The magnitude which 'electricity' was coined to refer to was given not by ostension, but by stating some of its manifest effects and, possibly, an elementary description of its causal powers, i.e., that, whatever else it is, electricity is capable of flow. The term 'electricity' could have been introduced on different occasions. In fact, André Ampère also introduced 'électricité' to account for a set of different effects, viz., currents and electromagnets. His own introduction was based on a different description, say, along the lines of a particulate electric fluid. What is there in common on all occasions where 'electricity' was introduced? What is there in common between Franklin's 'electricity', Ampère's 'électricité' and indeed, anybody else's 'electricity'? Putnam's 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" (1975a: 200).

Even if all these descriptions of electricity were wrong, according to the causal theory they would still be misdescriptions of electricity, rather than descriptions of nothing at all (cf. Putnam, op.cit.: 201). So, it seems to me that Putnam's response boils down to the following claim: what there is in common in all occurrences of the term 'electricity' in different theories and descriptions is that they all refer to the physical magnitude causally responsible for salient electrical phenomena. This physical magnitude is the referent of the term 'electricity' and guarantees the sameness in reference of the occurrences of this term. It then follows that in the case of the reference of theoretical terms, the "existentially given thing" is nothing but a causal agent, i.e., an agent which is posited to have the causal power to produce certain effects. A quick worry here might be that there is no guarantee that there is just one causal agent which brings about all these phenomena, e.g., electric currents, lightning bolts, deflections of magnets etc. This might well be so, but the causal theorist would quickly dismiss this worry by noting that he is not concerned with the epistemological problem of how we can come to assert that all these phenomena are due to electricity. All he is concerned with, he would point out, is show how all these different tokens of the term 'electricity' may nonetheless refer to the very same entity. On his view, if it happens that electricity is not responsible for, say, lightning bolts, then Franklin's 'electricity' does not refer to electricity, after all. But we have no reason to think that it does not.

A more promising critique is that given that the causal theory reduces referential stability to the bare assertion that a causally efficacious agent operates behind a set of phenomena, continuity and sameness in reference become very easily satisfiable. If the unobservable causal agent behind a set of phenomena is given only existentially, and if no description of its nature is essentially employed in fixing the reference of the terms that purports to refer to it, 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 critique negatively, on a pure causal account, it is not clear at all what could possibly show that the relevant theoretical term does not refer. If the reference of theoretical terms is fixed purely existentially, then insofar as there is a causal agent behind the relevant phenomena, the term is bound to end up referring to it. Hence, there can be no referential failure--even in cases where it is counter-intuitive to expect successful reference. Taken to its letter, the causal theory makes referential success necessary.

Take, for instance, the case of 'phlogiston'. It offers a neat example in which the pure causal theory yields counter-intuitive results. 'Phlogiston' was introduced on many occasions by means of a causal description, i.e., one that singled out phlogiston as the physical magnitude causally involved (given off) in combustion. Phlogiston, however, does not exist; hence, it is not causally involved in combustion. Instead, oxygen is. Does this mean, or imply, that phlogiston theorists had been referring to oxygen all along? If we follow the letter of the causal theory and accept that 'phlogiston' was coined to refer purely existentially to whatever is causally involved in combustion, then it is inescapable to conclude that 'phlogiston' refers to oxygen, since the latter is what is causally involved in combustion. This is surely far-fetched. Joseph Priestley and other advocates of the phlogiston theory were causally connected to oxygen--they breathed it and it was causally involved in the experiments they made to investigate combustion. But no properties of oxygen were the causal origin of the information they had associated with phlogiston. And nothing in nature could possible be the causal origin of such information. What is correct to say is that 'phlogiston' refers to nothing. But in order to say this, we need to say that there is nothing in nature which possesses the properties that phlogiston was supposed to possess. That is, we need to say that there is nothing which fits a description which assigns to phlogiston the properties it should have in order to play its intended causal role in combustion. So, once again, we have to rely on theoretical descriptions. 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 descriptions, and I content--by symmetry--so should be the case for referential success.

MORAL: Do we need to go back to descriptivism?

V. VERIFICATIONISM

Criterion of cognitive significance, i.e., criterion of having sense

"The meaning of a sentence consists in its method of verification".

- **The general project: circumscribe the boundaries of the meaningful**

a. science vs metaphysics
metaphysical statement violate strictures on meaning
surface structure of grammatical sentences, but really meaningless (pseudo-propositions)

Meaningful statements are of two sorts: either analytic (true in virtue of their meaning) or synthetic (true in virtue of facts).

All analytic statements are known a priori and only analytic statements are known a priori.

Synthetic (factual) statements are meaningful iff they are verifiable.

- **Cognitive Significance**

Verifiability as a criterion of cognitive significance

verification-conditions vs truth-conditions

the conditions under which a sentence can be asserted as true or false vs. the conditions under which a sentence is true or false. Sentences can have truth-conditions without having verification-conditions. And conversely, sentences can have verification-conditions which do not coincide with their truth-conditions. But, the verification-condition of a sentence can be an indicator of its truth-conditions.

Alternatively, one can take verification-condition to provide a reductive account of truth-conditions. Then, the truth-conditions of a sentence should be translatable into its verification-conditions.

Which of the two options did LP favour?

They started with the second. e.g., attempt to translate all statements into statements which refer to sense-data (and which are verifiable).
Then they move to observation sentences which have verification-conditions vis-à-vis theoretical sentences, but truth-conditions vis-à-vis sense-data sentences.
But the project comes to grief when it is accepted (circa 1950) that the evidence for the truth of a sentence should not be confused with the conditions under which the sentence is true (epistemological fallacy).
How are we to understand verifiability?

verifiability in practice vs verifiability in principle

strongly verifiable vs weakly verifiable

strong: its truth can be conclusively established in experience
weak: its truth can be rendered probable by experience.

All these will give different formulation of the criterion of meaningfulness. Some sentences which are meaningful on one version might not be meaningful on another.

The most liberal version is weak in principle verifiability. But this is not, strictly speaking, verifiability. It's confirmability. And this is consistent with the sentence having truth-condition and being meaningful because it has truth-conditions.

Verifiability is neither necessary nor sufficient for meaningfulness.

Unnecessary: statements that report laws of nature are meaningful and yet not verifiable.

Insufficient: Take meaningful sentence O and meaningless sentence N ("the absolute is sleepy"). O entails (O or N), and hence it verifies (O or N). But (O or N) is hardly meaningful.

What about the criterion of meaningfulness itself? Is it verifiable?

▫ Analytic vs synthetic--A priori vs A posteriori

What do they mean?

Analytic vs synthetic: what makes a sentence true if it is true.
weaker version: how truth ascription is justified. (grasp of meaning alone suffices for justified belief in its truth)

➢ Kant's two accounts of analyticity:

i) the meaning of the predicate is contained in the meaning of the subject.

So: facts about meanings. This is not innocent. It reflects a certain metaphysics of essential and accidental properties which was characteristic of the Aristotelian logic.

e.g., "Man is rational animal".

weaker: facts about how meanings are used.

ii) a sentence is analytic if it's negation is not contradictory (or if it can be derived from the law of non-contradiction)

Given Aristotelian Logic, arithmetical sentences cannot be analytic.

Criterion (ii) is the most important. But, given Aristotelian Logic, (i) and (ii) coincide.

A priori vs A posteriori: how is the truth of a sentence known?

stronger version: what is the modal status of the sentence (necessary or contingent)?

Kant goes for the stronger version. "Necessity and universality are sure criteria of a priori knowledge" (B4).

Kant: there is a body of synthetic a priori truths (e.g., arithmetic, geometry, general principles of science, e.g., the causal maxim)

	Analytic	**Synthetic**
A Priori	YES	YES
A Posteriori	NO	YES

Why synthetic? Because they are not analytic.
Why are they a priori? Because they are not a posteriori.
Why are they not a posteriori? Because they are necessary.

Kantian thought: Necessity is a sure route to a priority. But the necessity of a judgement does not dictate its content. The judgement can be either synthetic or analytic. So, there can be both analytic a priori and synthetic a priori judgements.

Sum up: Kant on A Priori Knowledge

a. Knowledge which is universal, necessary and certain
b. yet it is contentful (if it is synthetic) or empty (if it is analytic)
c. its content is formal: conceptual connections (if analytic); form of pure intuition (if synthetic)
d. constitutive of experience (no experience is possible without it)
e. disconnected from the content of experience (although constitutive of its form). Hence unrevisable.

So, synthetic a priori knowledge is contentful and yet necessary and universal.
It's being contentful is guaranteed by being about pure intuition and its necessity is guaranteed by being about the form of pure intuition.

pure intuition as a source of contentful yet necessary and universal judgements.

empirical intuition vs pure intuition

empirical concepts can be amplified synthetically in empirical intuition, e.g., all ravens are black.

Pure intuition: "as intuition a priori it is inseparably joined with the concept before all experience or particular perception" (Prolegomena). So, some synthetic judgements are a priori certain and apodeictic.

Both "5+7=12" and "Man is rational animal" are a priori true. Yet, "5+7=12" is not analytic. Hence it is synthetic. And "Man is rational animal" is analytic.

Kant's argument from the large numbers.

33542+18349=51891

Kant's Moral: "analytic" and "a priori" are not co-extensive concepts.

Important issue: what makes a priori truth true? For Kant it cannot just be facts about meanings, since there can be synthetic a priori truths. So, it seems that what makes them true are facts about their modal status, viz., that they are necessary. Yet, Kant does not have a linguistic doctrine of necessity: not just analytic truths are necessary truths. But what exactly is his sense of necessity?

- **Frege: Partial Rejection of the Kantian synthetic A priori**

Frege rejects Aristotelian Logic (AL). Hence, he rejects the first Kantian criterion of analyticity. This criterion is sufficient for analyticity, but not necessary: a sentence can be analytic even if the concept of the predicate is not contained in the concept of the subject. "5+7=12" is analytic for Frege. For Frege the second Kantian criterion is enough for analyticity, but the second does not imply the first if AL is rejected.

Arithmetic is analytic, because arithmetic reduces to logic (but N.B. this is the New Logic).
But for Frege, geometry is synthetic a priori.

	Analytic	**Synthetic**
A Priori	YES	YES
A Posteriori	NO	YES

Central notion: proof from primitive truths.

Frege-analyticity: a sentence is analytic if it is derived from general logical laws and definitions.
So, a sentence is synthetic if its proof requires non-logical truths.
Frege-a priority: a sentence is a priori if its proof depends only on general laws which need no and admit of no proof.
So, a sentence can be a priori without being analytic (e.g., geometrical truths).
Besides, analyticity is something that gets transferred from the truths of logic and definitions, via a derivation, to another sentence that logically follows from them.
But what exactly gets transferred? What exactly is it that logical truths have that makes them analytic?

Three thoughts: a) they satisfy the law of non-contradiction. But this is another logical law.
b) they are universally applicable ("universal scope"). Their domain is "everything thinkable". They have no special subject-matter.

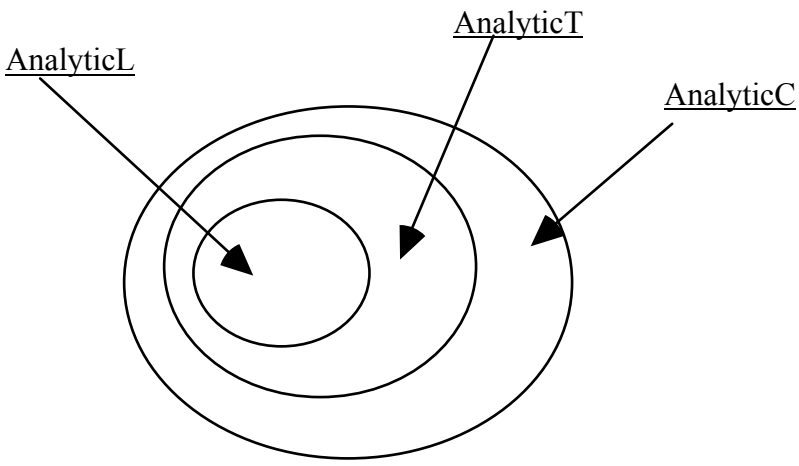
Frege's hierarchical view of knowledge.
synthetic a posteriori truths (sciences)
synthetic a priori truths (geometry): spatial intuition is required (synthetic); they govern "all that is spatially intuitable" (a priori)
Both of these truths can be denied without contradiction. Hence they cannot be derived form logic alone (+ definitions).
analytic a priori truths (logic and arithmetic): universal scope.

c) they need no intuition; proof is enough. But the very general principles from which they derive are not provable. Still, their truth cannot be denied, so back to (a).

Important issue 1: what is it that makes the most general truths of logic analytic? We can say what they are, what we can do with them, but what do we ascribe to them when we say that they are analytic?

Important issue 2: are there analytic statements which are not logically true? This is a hot issue. Analytic_L: the logical truths. Analytic_T: a broader class of statements which are true in virtue of their meaning. Kant tried to characterised them by his first criterion. But can we characterise them if we abandon Aristotelian Logic? We need facts about synonymy. Are there such facts, what do they state and how do we know them? Frege-analyticity: **a sentence is analytic if its is transformed into a logical truth by the substitution of synonyms for synonyms**.

Suppose that Frege-analyticity explains why some sentences are analytic. Hence it also explains why they are also a priori, since they are reducible to logical truths and these are a priori. Two more problems arise. First, this move leaves the analytic and a priori status of logical truth unexplained, since it presupposes that logical truths are analytic and a priori. Second, it leaves unaccounted for a number of statements which appear to be a priori true and yet do not satisfy Frege-analyticity. E.g., "Nothing can be red and blue all over at the same time", or "If x is warmer than y and y is warmer than z, then x is warmer than z". Can they be seen as Analytic_C?



Important issue 3: what principles belong to logic? So what principles count as Analytic_L?

Propositional calculus: tautologies (the easiest case for analyticity)
They are decideable and empty of content. But no distinction between Analytic_L and Analytic_T can be drawn within the prop calculus.
Set-theory: but set-theoretic paradoxes.
Predicate calculus: problems with infinity.

Frege's Moral: "analytic" and "a priori" are not co-extensive concepts. Yet, logic and arithmetic are analytic a priori.

- **Revision of the Kantian synthetic a priori**

The early Reichenbach

Kant's a priori: a) necessarily true
b) constituting the object of knowledge.

Reichenbach rejects the first account of a priori, but insists that the second is inescapable.

Principles of co-ordination are "constitutive principles of experience".

co-ordination: how basic concept are connected to reality.

Hilbert-style implicit definition and their role in maths (here only consistency is required).
Generalisation of Hilbert's programme to physical theories. Here we need more than consistency--we need empirical content. This is effected, partly, by means of co-ordinationg the basic concepts with the world.

axioms of connection vs axioms of co-ordination

axioms of connection (empirical axioms): connecting certain state-variables
axioms of co-ordination: general rules according to which connections take place (e.g., the rules embodied in the relevant maths). These "precede the most general laws of connection". They are a priori in the sense that the constitute the object of knowledge.

Important issue: This may be a useful but thin notion of a priori. All it point at is a functional distinction between two kinds of axioms. The co-ordinating axioms have a special logical function in the construction of knowledge. But they are not different from the axioms of connection in any essential sense: e.g., they are both synthetic and revisable. In fact, Reichenbach takes Kant's mistake to be that he thought the co-ordinationg axioms to be necessary and unrevisable. Or is the difference bigger?

Reichenbach's argument against Kant.
Kant thought that the experience cannot lead to the revision of the co-ordinating principles because these principles are presupposed by experience. So, in effect Kant thought that there is one unique co-ordination between these principles and experience. But, R. says, these principles may overdetermine experience. So, there is no guarantee that these principles allow for a unique co-ordination with experience. And there is no guarantee because for Kant these principles originate in reason and perceptions (which provides the criterion for uniqueness of co-ordination) is independent of reason.

In fact, R. aims to show that some co-ordinating systems may prove to be incompatible with experience.

The relevance of the Duhem-Quine thesis. Can't any co-ordinating system be saved from refutation?
e.g., Euclidean Geometry + Strange Physics = Non Euclidean Geometry and Normal Physics.

Yet, the first option (which includes universal forces) does not allow for unique co-ordination. Whereas a rigid-body hypothesis does.
Co-ordinating principles are not subject to confirmation and disconfirmation. But they make possible the conformation/disconfirmation of empirical laws. Yet, what accounts for their own revisability?
The irrelevance of the Duhem-Quine thesis: Schlick vs Reichenbach. For Schlick, who also spoke of constitutive principles, the difference between constitutive principles and others is epistemic. When we test a theory we test it as a whole, so, for instance we test Geometry and Physics together, and therefore we can choose what to revise (Physics or Geometry) based on considerations of simplicity. On this view, there is no big difference between constitutive principles (e.g., Geometry) and other principles (e.g., Physics). This amounts to Quinean holism. But Reichenbach's aim is different. He wants to argue that although both kinds of principle are revisable, there is still a logical difference between them. The constitutive principles mark off the non-empirical, whereas the axioms of connection mark off the empirical. Still, it's not clear how they do that and what the difference is.
Note that for Reichenbach the co-ordinating principles are revisable. They do not come temporally prior to the axioms of connections. He says: these principles can be "detected only gradually by means of logical analysis" (pp.77-78).

- **Full Rejection of the Kantian synthetic a priori**

  - **Logical Empiricism**

Completing Frege's project.
Geometry is not synthetic A priori
revisability, but shared premise: a priori=necessary (all necessity is linguistic)
rejection of spatial intuition: Hilbert's arithmetisation of geometry)
So, all a priori statements are analytic.

	Analytic	**Synthetic**
A Priori	YES	NO
A Posteriori	NO	YES

But what are analytic statements? **Truths by definition or truths by convention**.

Demystifying analyticity and a priority. No special faculty of intuition; no epistemic problems in accessing them.

N.B. LP accept Kant's second criterion of analyticity and Frege's only criterion. The denial of an analytic truth would be "self-stultifying" (Ayer). But they seem to offer an explanation of it, viz., that in denying an analytic statement we would violate some basic rules of language.

For instance, the law of non-contradiction expresses a linguistic prescription of how to speak about things.
Analytic truths are true by virtue of definitions or of similar stipulations which determine the meanings of the key terms involved in them.
So, there are facts about synonymy, but these facts are essentially conventions about the meanings of words.

"7+5" is synonymous to "12".

Frege-analyticity works because there are (conventional) facts about synonymy.
Besides, statements in the class Analytic_C are indeed analytic, since they do not report facts about the world, but facts about the use of relevant predicates, e.g., the colour predicates.

Analytic truths are factually empty--they have no empirical content.
Analytic statements are constitutive of a language.

Linguistic doctrine of necessity. Analytic truths (and only those) are necessary.
This is supported by two considerations: First, they are factually empty--they have no empirical content, so their content (and truth-status) cannot vary. Second, (an epistemological consideration) they appear to be genuinely unrevisable--they cannot be confirmed (this is especially apparent in Ayer).

Carnap: **meaning postulates**

L-truth: logical truths are satisfied by all values of the descriptive constants that occur in them.
A-truth: meaning postulates
They assert certain logical connections between predicates. e.g., "All Bachelors are unmarried men".
It's a matter of decision what sentences will deem meaning postulates.

The Logical Empiricists take analytic statements to be necessary. Hence they partly agree with Kant. They take them to be constitutive of experience, since they specify the meanings of concept we use. So, they also partly agree with Kant. They also state that analytic statements are a priori and hence they accept the existence of a priori knowledge. So, their fundamental disagreement is about the synthetic a priori knowledge.
Analytic statements are neither confirmable nor disconfirmable. Yet, the later Carnap especially insists that they are revisable. How can that be? Carnap offers a language-specific analysis of analyticity. So, new language, new analytic statements.

Quine's fundamental objection: Are there truths by convention? In particular, can logical truths be true by convention?

A and B	A and B	A, B
---	---	---
B	A	A and B

These schemas cannot define the meaning of "and" by convention, for there are infinite instances of them. So, we need an explicit formulation of the form "all instance of the schema with premises 'A and B' and conclusion 'A' are valid". But clearly this sentence has an occurrence of 'and' in it and so it cannot define the logical 'and'."

Quine: "the difficulty is that if logic is to proceed mediately from the conventions, logic is needed for inferring logic from conventions".

- **Quine**

Quine completes the critique of Kant by rejecting the existence of analytic truths. So, since the LP had already rejected the existence of synthetic a priori truths, Quine's rejection of the analytic a priori means that he rejects a priority altogether.

Quine on "Two Dogmas of Empiricism"

Analytic_L: logical truth (coincides with Carnap's L-truth above).
Analytic_T: truth in virtue of meaning only.

First Objection: Analytic_T requires a notion of cognitive synonymy. Yet there is no independent criterion of cognitive synonymy.

In particular, all attempts to characterise cognitive synonymy will involve the notion of analyticity.
co-extensionality (substitutivity salva veritate) is a necessary condition for synonymy but not sufficient. So, no criteria for sameness of sense.
Grace and Strawson's reaction: absence of conditions for synonymy implies absence of meanings. But Quine accepts this. But he offers an empiricist approach to meaning and language in which meanings (and senses in particular) are not reified.

Why not hold on to Analytic_L?

Quine's second objection: "Analytic_L" does have a definite extension (i.e., the truths of logic). But what is it ascribed to them when they are called analytic? No definite property that can be circumscribed by the term "analytic".

Quine's third objection: "Analytic" is taken to mean "unrevisable". Since analytic statements have no empirical content, experience cannot possibly have any bearing on

their truth-value. So analytic statements can suffer no truth-value revision. But nothing is unrevisable. So, there cannot be any analytic truths (either Analytic_L or Analytic_T).

What is the argument for the revisability of all statements (beliefs)?

Confirmation and refutation are both holistic. So, anything (i.e., any sentence) can be confirmed (so, everything has empirical content). And anything can be revised (so, everything can be such that experience leads to revision of its true-value).
With this argument Quine shows that there cannot be analytic a priori. If there are no analytic statements, then (given that LP showed that there are no synthetic a priori statements), there are no a priori statements.

If analytic then unrevisable.
Nothing is unrevisable.
Therefore, nothing is analytic.
Therefore nothing is a priori.

Carnap disagrees with Quine on the major premise. He thinks that analyticity does not imply unrevisability. So, for Carnap there is no relation between the two concept and therefore Quine's argument fails.
How successful is Quine's argument?

For Carnap analytic statements are such that a) it is rational to accept them within a linguistic framework; b) rational to reject them subsequently, when the framework changes; and c) there is some extra characteristic which all and only analytic statements share.
Note that Carnap needs (c) because (a) and (b) alone do not mark out the difference between analytic and synthetic statements. So even if we grant that analytic statements are revisable, we still need a way to characterise how they are different from synthetic statements (within a framework). So, Quine's first objection still stands: to just call them "analytic" casts no light on their supposed special status.

Quine's fourth objection: if confirmation is holistic then each particular statement cannot have its own conditions of confirmation. Hence, insofar as meaning is tied to confirmation, then statements cannot have meaning when taken individually. Only whole systems of statements can be synonymous (insofar as they have the same conditions of conformation). Enter indeterminacy of translation. No two individual sentences can be taken to be synonymous. Hence, there are no facts about synonymy and hence there are no facts about meanings.

	Analytic	**Synthetic**
A Priori	NO	NO
A Posteriori	NO	YES

Carnap on External/Internal Questions

In his "Empiricism, Semantics and Ontology" (1950 [1956]), Carnap attempts to disperse the view that talking about and quantifying over certain entities implies a metaphysical commitment to them, a commitment to their independent existence. His suggestion is that questions of existence of a certain kind of entities can be understood in two different ways: either as external questions or as internal ones. External questions are meant to be metaphysical in nature: they concern the existence or reality of "the system of entities as a whole" (ESO, p.206). Answering questions such as 'Are there entities of such and such nature? (e.g., numbers, properties, classes etc.)?' is taken to presuppose that the existence of such entities can be asserted or denied independently of a certain discourse. Qua external, such questions require that one must first establish the existence of such entities before one starts talking about them. For Carnap all this is fundamentally wrong. No metaphysical insight into their nature is needed for the introduction of a new kind of entities. Instead, all that is needed is the adoption/construction of a certain linguistic framework whose linguistic resources makes it possible to talk about such entities. Once the framework is adopted, questions about the existence or reality of the relevant entities lose any apparent metaphysical significance. They become internal: it follows from the very adoption of the framework that such entities exist. Hence, asserting their existence becomes analytic. On this way of looking at things, what remains of the external questions relates to a

certain practical decision to adopt the certain linguistic framework. No facts in the world will force us adopt a certain framework—least of all the alleged independent existence or reality of the talked about entities. The only relevant considerations are pragmatic: the efficiency, fruitfulness and simplicity of each proposed linguistic framework. Take, for instance, questions of the existence of space-time points. Seen as external questions, they are either metaphysical pseudo-questions ('Are there (really) space-time points?) or practical questions concerning the adoption of framework for the development of scientific theories. In the latter case, once the framework has been adopted, it's an analytic truth that there are space-time points (cf. ESO, pp.212-213).

What exactly does it mean to introduce a new framework, or as Carnap put it, a system of entities? He specifies two conditions. 1) The introduction of a new general term T (e.g., 'number', 'property', 'proposition' etc.) for the new kind of entities, so that one can make statements of the form 'ϕ is T'. 2) The introduction of a new type of variable that ranges over the new entities, so that one can make statements of the sort '$\forall\phi$ if ϕ is so-and so, then ...' (cf. 1950, pp.213-214). After the new general term and the new type of variable have been introduced in the framework, we can ask internal questions about these entities, which admit of either analytic or synthetic answers. But as Quine (1951) has convincingly shown, the whole distinction is suspect. His main point is this. The external/internal distinction is a special case of a putative distinction between category questions and subclass questions. The former relate to the admissibility of a new category of entities, e.g., numbers, or propositions, while the latter relate to the admissibility of certain subclasses of such entities, e.g., prime numbers, or equivalent propositions. Seen that way, the distinction is ill-conceived. In order to separate the two types of question and treat their existential implications differently, we need to introduce segregated styles of variables, e.g., variables that range exclusively over numbers without ranging over subclasses of them. But there is no need, nor motivation, to do such thing. This might make sense from the point of view of the Russellian theory of types. But even there, and more so in set theory, it is not necessary: categories and subclasses of them do not necessitate different styles of quantification.

Carnap is quite right in stressing that “The purposes for which the language is intended to be used, for instance the purpose of communicating factual knowledge,

will determine which factors are relevant for the decision" (1950 [1956], p.208). One can easily accept that, in the true spirit of the Principle of Tolerance, there are no pre-set aims for the construction of a language. But if the aim is as in Carnap's example, i.e., the communication of factual knowledge, then the next question is this: do we have the luxury of choosing any other language than one that employs theoretical terms and implies commitments to theoretical entities? In particular, do we have the luxury to employ a language that involves only observation terms? Let's grant that in devising the language of science we want to communicate factual knowledge. If this cannot be achieved in an observation language, if that is, it requires the employment of a theoretical language, why aren't all these considerations of simplicity, fruitfulness and expedience justification enough for the adoption of the theoretical language and of whatever existential commitments it implies?

VI. QUINE'S MEANING SCEPTICISM

- **The argument of "Two Dogmas"**. No clear conditions of meaning-identity (synonymy).

- **The argument from the Indeterminacy of Translation.**

Thesis: There is no fact of the matter as to what an expression means. All there is to meaning is dispositions to assent to or dissent from certain assertions. These dispositions underdetermine the content of the assertions. Yet, there is nothing further can make this content determinate.

Radical thesis

In more detail:

Meaning is linked to translation. To say that expressions E1 and E2 have the same meaning is to say that one is the translation of the other. Take languages L1 and L2. A translation manual is a scheme which translates expressions of L1 into expressions of L2. Intuitively, there is a unique correct translation manual: the one that pairs expressions of L1 with those expressions of L2 that have the same meaning. Quine's thesis: there is no uniquely correct translation manual. In fact, there will be an infinitely large number of mutually incompatible translation manuals which fit equally well all of the possible evidence relevant to translation. So, there is no fact of the matter as to which is the correct manual.
N.B. "all possible evidence relevant to translation"
Quine rejects the museum theory of meaning. There is no direct apprehension of senses. All evidence should be empirical. Ultimately, it should relate to dispositions to behaviour.

Two arguments for this thesis.

- **The argument from below.**

Radical translation

Why take radical translation seriously? a) Ordinary translation takes for granted lots of disputed things, e.g., assumptions about what we mean, what dictionaries say etc. b) Radical translation brings to focus the link between meaning and behaviour: only behavioural facts are present in radical translation. c) any kind of translation should be accessible form the vantage point of radical translation. If there is a correct manual between, say, Greek and French, both us and a Martian should be able to formulate it. There are no extra facts that we may get and a Martian miss.
The only relevant evidence is behavioural--in a narrow sense: there are no semantic or intentional facts to be taken into account. Only stimulus meaning is relevant to radical translation.

Stimulus meaning: <sensory stimulations that prompt assent to the sentence, sensory stimulations that prompt dissent to the sentence>

Stimulus synonymy: two expressions are stimulus synonymous iff they have the same stimulus meaning.

Thesis: in radical translation (and hence in all translation), there is no further fact of the matter as to which of the many possible stimulus-synonymous expressions is the correct translation of a certain native expression

e.g., "gavagai"

There goes a rabbit.
There is an undetached rabbit part

etc.

All are stimulus-synonymous. No fact of the matter as to which is the correct rendering of "gavagai".

N.B. Quine is not an epistemic sceptic. He does not say that there is a fact of the matter as to what a native expression means, but we don't (can't) know it. He's a radical sceptic. There is no such fact of the matter. There is no such thing as THE meaning of a native expression.
How does radical translation proceed? By means of working hypotheses. e.g., affirmation and negation. (W&O, pp. 29-30)
How do we choose the translation? Not based on some facts about meanings but on pragmatic grounds.
Can we eliminate alternative manuals by taking the translation of other worlds into consideration? Not really. The indeterminacy extends to all words in the language. (e.g., principles of individuation).

Evans' answer to Quine

Distinguish between theory of translation and theory of meaning. Even if translation is indeterminate, it does not follow that there is no fact of the matter as to what a speaker means by a word (or expression).

What is a theory of meaning for a language L?

1) A systematic theory which states what the meanings of expressions of L are. In particular, it is a compositional theory which analyses the meanings of expressions in terms of the semantic values of their constituents.

e.g., Frege's truth-conditional theory of meaning.

2) Mirror Constraint: the theory of meaning should be such that it mirrors the linguistic abilities of the actual speakers of L.
These abilities are such that a normal speaker of L is able to move to an understanding of a sentence S by understanding its constituents as they appear in other sentences than S.

Important factual claim behind the Mirror Constraint: new sentences are understood because of grammatical structure + familiarity with the meaning of their constituent words.
How is this an argument against Quine's scepticism about meanings? A theory of radical translation fails to establish a meaning theory for the native language. And a meaning theory (with the mirror constraint) does not fall prey to indeterminacy).

Example:

Suppose we set out to construct a theory of meaning for the native language. We need to specify the meanings of terms such as "gavagai" and "blap". Hypothesis: "gavagai"--->rabbit; "blap"--->white.

Axioms in the theory of meaning:

(1) For all x (x satisfies "gavagai" iff x is a rabbit).
(2) For all x (x satisfies "blap" iff x is white).

Given this we can specify the truth-conditions for the native expression "gavagai blap": rabbits are white.

Quine would argue that another axiom (1') can replace (1) in a different translation manual without any conflict with the native's behavioural dispositions.

(1') For all x (x satisfies "gavagai" iff x is an undetached part of a rabbit).

But (1') will give the wrong truth-conditions for other complex expressions of the native language in which "gavagai" occurs. Example: "gavagai blap".

On the basis of (1) and (2) we can construct the truth-conditions of "gavagai blap".

(3) "gavagai blap" is true iff a is a white rabbit.

But on the basis of (1') and (2) we can't get the right truth-conditions

(4) "gavagai blap" is true iff a is a white undetached part of a rabbit.

This is not an adequate rendering of the truth-conditions because it would not be assented to by the natives. (4) can be assented to in the presence of a brown rabbit with white foot. But natives wouldn't assent to it.

Quine's move: Revise (2) too.

(2') For all x (x satisfies "blap" iff x is a part of a white animal).

(1') and (2') specify the truth-conditions of "gavagai blap" as follows:

(3') "gavagai blap" is true iff a is an undetached rabbit part of white animal.

The natives would assent to it in the presence of a white rabbit, but wouldn't assent to it in the presence of a brown rabbit with a white foot. Yet, (3') violates the Mirror Constraint. If the meaning of "blap" is specified as described by (3'), then it cannot function properly when it is conjoined with other words:, e.g., when it describes a piece of paper.

Hookway's manoeuvre:

(4) For all x (x satisfies "blap" iff either (a) "blap" occurs together with "gavagai" and x is part of a white animal, or (b) "blap" occurs in some other context and x is white).

How many dispositions to assent?

- **Quine's Argument from Above**

This is a different argument. It rests on the underdetermination of theories by evidence, but it goes much further than this.

"The real ground for the doctrine is very different, broader and deeper" (1970, 178).

Radical translation of a foreign physicists' theory.

First step: Start as above. At the level of observations (observational consequences). Stimulus synonymy. Setting up of a translation manual (by means of analytical hypotheses).
Second step: Try to translate the foreign theory into ours. Here is where underdetermination strikes us. Even if we fix the translation of the observation sentences "the translation of the foreigner's physical theory is underdetermined by translation of his observation sentences" (1970, 179).

Suppose O_f is the observational sentences of the foreign theory T_{0f}. Given underdetermination, there will be a number of theories T_{1f}, ..., T_{nf} which are empirically equivalent with respect to O_f. Suppose (first step) that we translate O_f into our own O_n. O_n underdetermines theories, hence there will be a number of our own theories T_{1o}, ..., T_{no} which are empirically equivalent with respect to O_n. So, (second step) it is obvious that we are free choose any of the T_{if}s into any of the T_{io}s.
Important: Quine's claim is that there is no further fact of the matter which would render one choice of theories (among the T_{if}s and the T_{io}s) THE correct one.
Also: Quine's indeterminacy thesis is "additional" to the problem of underdetermination. Quine does not doubt that there is a true theory of the world. But he does doubt that there are facts about meanings that make one translation THE correct one.
The totality of physical truths determines the correct theory of the world. But it does not determine facts about meanings, and hence the correct translation.
But is Quine's thesis cogent? It all depends on the nature (and scope) of underdetermination.

VII. MEANING AND TRUTH-CONDITIONS

Truth-Conditional Semantics (Davidson)

To know the meaning of a sentence is to know its truth-conditions

what are the truth-conditions?
distinguish from verification- or evidence-conditions

Synonymy and truth-conditions

Entailment (logical relations) and truth

1) A theory of meaning should help us understand what determines the meaning of a particular sentence
2) A theory of meaning should explain what the understanding of sentences (especially novel sentences) consists in.

Re (1): A theory of meaning for a particular natural language, e.g., English or Greek

Compositionality

The meaning of a sentence is a function of the meanings of its constituent words

This helps with (2).

So, we start with individual words "atoms" (linguists call them "moprhemes") and compute the meaning of a sentence by syntax + atoms.

- **Theory of meaning for a native language**

A sequence of sentences (theorems) that specify the meanings of the sentences of the language

e.g., "Snow is white" means that snow is white
"Grass is green" iff grass is green
etc.

Then, the task is to specify the meanings of the sentences in terms of their truth-conditions.

"Snow is white" is true iff snow is white
"Grass is green" iff grass is green
etc.

Here Tarski's schema (T) is very helpful.

'------' is true-in-L iff ------

Instances of this schema can specify the meaning of sentences '-----' by specifying the conditions under which they are true. What else would we need for a theory of meaning?
N.B. Davidson does not want to reify meanings. Meanings (senses) are not abstract entities. So, he is not a Fregean—although he takes seriously the Fregean view that meaning is linked to truth-conditions. In a sense, Davidson offers a referential theory of meaning.

Compositionlality for truth-conditions

Complex (molecular sentences): their truth conditions are a function of the truth-conditions of their atomic constituents
Atomic sentences: their truth-conditions are a function of the semantic values of their constituent expressions

- **A foreign language**

"Το χιόνι είναι άσπρο" means [in Greek] that snow is white

How is this translation possible?

The two sentences share the same truth-conditions. The truth-conditions of the Greek sentence "Το χιόνι είναι άσπρο" are specified as follows:

"Το χιόνι είναι άσπρο" is true (in Greek) iff snow is white. The sentence used in the right-hand side specifies the truth-conditions of the sentence mentioned in the left-hand side.

Sameness of meaning is then sameness of truth-conditions

- **The first-order predicate calculus seems to offer a nice model for the semantics of natural languages**

Truth-definition for a language

From a finite number of rules (steps) we can determine the truth-conditions (and hence the meanings) of an indefinite number of sentences.

➢ Tarski's definition of Truth

Languages, especially formal languages, are usually employed to describe structures. A structure Υ is defined as follows:

$$\Upsilon=(U, I)$$

where U is a non-empty set of objects or individuals, which is called the universe of discourse or domain, and I is an interpretation function which makes the following mappings:

(1) To every term t of L, I assigns an object u of U. That is, $t^I=u$.
(2) To every n-ary function symbol f of L, I assigns an n-ary operation on U.
(3) To every n-place predicate P of L, I assigns an n-ary relation R on U. That is, $P^I=R$. In other words, I assigns an n-tuple of objects of U as the extension of the n-place predicate P.

These mappings provide an L-interpretation of the structure Υ. Besides, all variables of L are assigned values from the domain U, such that for every variable v of L, $v^I=y$, where $y \in U$. This procedures is known as L-valuation or pretence. We will adopt the term valuation, and in particular σ-valuation, for an assignment of values to the variables of L whose other symbols are already interpreted by the interpretation function I. Therefore, σ-valuation agrees with Υ over the domain and the interpretation function I. (We can think of σ as an extension of the interpretation function I.) Then we shall say that a valuation σ-satisfies a sentential function (or an open sentence/formula) ϕx iff $(\phi x)^\sigma$ is true under this valuation. In particular, if a sentence ϕ has no free variables we shall say that the underlying structure Υ is a

model for ϕ if ϕ is true in this model. For example, the sentential function (Fx), where 'F' is interpreted as 'Professor of Philosophy in Athens, is satisfied by x^{σ}, iff the value of x under valuation σ belongs to the extension of F; or, equivalently, iff a true sentence results when we replace x by its value fixed by the specific valuation. That is, if x^{σ} is Anapolitanos then (Fx) is clearly true. In this sense, a (Tarskian) atomic formula (open sentence) is thoroughly referential, even though it is not a closed sentence. For under this particular value assignment, all variables behave as if they were singular terms with objects assigned as their referents. Hence the notion of reference, as it is captured by both the interpretation function and the valuation, becomes indispensable in Tarskian semantics (cf. Davidson, [1990], p.296). Given that we have fixed the model of L, that is given that we have seen L as representing faithfully a structure, we may lay down the recursive definition of truth-in-L, à la Tarski. This is accomplished in two steps:
(1) Every atomic formula (i.e. sentential function/open sentence) is mapped onto the set {true, false} according to the following condition :

$(Pt1...tn)^{\sigma}$ is true iff $<t1^{\sigma},...,tn^{\sigma}> \in P^{\sigma}$,

which says that an n-tuple of terms <t1,t2,...,tn> under a given valuation σ satisfies an atomic formula $(Pt1...tn)^{\sigma}$, where P is an extra-logical n-place predicate of L, iff <t1,...,tn> under this interpretation/valuation belongs to the extension of P under the specific interpretation. This is nothing but another way to state the intuitive convention (schema) T. If P is monadic for instance, it amounts to stating that, for instance, "snow is white" is true iff snow is white.

(2) For all atomic formulas ϕ, ψ,
(i) $(\text{not-}\phi)^{\sigma}$ is true iff ϕ^{σ} is false.
(ii) $(\phi \ \& \ \psi)^{\sigma}$ is true iff ϕ^{σ} and ψ^{σ} are true.
(iii) $(\forall x\phi)^{\sigma}$ is true iff $\phi^{\sigma(x/u)}$ is true for every u □ U. ($\sigma(x/u)$ says that x takes the value u in the valuation σ.)

Clauses (2) of the definition are the so-called recursive clauses. They define the truth-values of molecular sentences in terms of the truth-values of their components. Moreover, they guarantee that the initial valuation is extended into a truth-valuation, that is, a valuation which preserves truth. It turns out that, in this definition, a sentence (i.e., a sentence with no free variables) is true if it is satisfied by any sequence of objects whereas a sentence is false if it satisfied by no sequence of objects. It is easy to see why this is so. A sentence has no free variables. Hence, the

satisfaction of this sentence does not depend on the values assigned to variables, since there are no variables to be given values. Hence, trivially, either every sequence of objects satisfies a sentence or none does. In the former case, the sentence is true, while in the latter it is false (cf. Tarski, [1933/1956], p.194). But, in fact, we may take this suggestion to mean the following: since a sentence contains no free variables, its truth-value is determined only by the interpretation of L and not by the additional assignment of values to variables. Since the interpretation is fixed in advance, any sentence comes out either true or false by, just, looking at the model.

Summary:

Tarski opted for the well-known *recursive* introduction of the truth-predicate by means of operations on atomic formulas—i.e., sentential functions which, in their simplest form, contain a monadic predicate followed by a (free) variable, viz. Fx. Sentential functions are not, however, either true or false. In order for these categories to apply, the sentential functions must be replaced by sentences, that is, their free variables must get values. Then, if, for instance, the variable x in the sentential function Fx gets object **a** as its value, one can say that the resulting sentence Fa is true iff **a** satisfies F and false otherwise; that is,

$$\mathrm{F}\mathbf{a} = \begin{cases} \text{true} \Leftrightarrow \mathbf{a} \text{ satisfies } \mathbf{F} \\ \text{false} \Leftrightarrow \text{otherwise.} \end{cases}$$

In other words, Tarski introduced the truth-predicate by means of the notion of *satisfaction*, which, according to him, can be rigorously defined (cf. 1933, 189 & 1969, 69). It is easily observed that the notion of satisfaction is akin to that of reference. The value that is assigned to a variable is the object referred to by the variable, under a particular value-assignment; the values of individual constants (closed terms) are the objects that these terms denote; and so forth for complex terms. Concerning atomic formulas with monadic predicates, the objects that satisfy the sentential function are no other than the objects that fall into the extension of the predicate featuring in this function, and so forth for more complex atomic formulas. That is, the truth of sentences depends on the semantic values (i.e., reference) of certain parts of the sentences (cf. Dummett, 1982, 59-60; Davidson, 1990, 296).

Some important issues

1. Are natural language like formal logical languages?

2. Even if truth-conditional semantics offers a theory of meaning for declarative sentences, do they manage to offer semantics for other types of sentences (e.g., commands, questions etc.

a. Such sentences may be deemed 'true' or 'false' in a non-standard sense (e.g. a command is 'true' if it is obeyed. But this is problematic.
b. compliance-conditions. This is better
c. Factually defective discourse (e.g., emotivism about ethics)
Two options: a) it is consistent to embrace both truth-conditional semantics and be an emotovist about some areas of discourse. b) one can use truth-conditional semantic to show that a certain discourse cannot be factually defective (e.g., the role of conditionals in ethical discourse).

3. Indexical expressions. Can they have Davidsonian truth-conditions?

"I'm sick now".
"I'm sick now" is true iff I'm sick now.
But this fails to capture the meaning of "I'm sick now", since its meaning will vary with the utterer and the place. (It will vary with the context of utterance.)
Davidson: relativise these sentences.
"I'm sick now" as it's uttered by p at t is true iff p is sick at t.

4. Not truth-functional connectives

e.g., because
intensional contexts
x believes that y believes that z believs that the food is great.

5. How exactly are the truth-conditions of a sentence specified?

This is an important objection. Two (related) types of counterexample.
Take two co-extensional expressions that are different in meaning.
e.g., "renates": animals with kidneys
"cordates": animals with heart.
Take sentence: "Renates have legs". Now, it appears that the truth-conditions of this statement can be thus: cordates have legs. Both sentences are true under exactly the same circumstances and hence they have exactly the same truth-conditions. But we rightly feel that the meaning of "Renates have legs" is different from the meaning of "Cordates have legs". So, there must be something other than their truth-conditions that capture their meanings. Hence, a truth-conditional account of meaning cannot capture properly the meaning of some sentences.
Take the sentence: "Snow is white". The truth-conditions of this sentence are exactly the same as the truth-conditions of the sentence "Snow is white and 2+3=5". Whenever the ones obtain so do the others, and conversely. (Note that the latter sentence is a conjunction). But clearly their meanings are different. So, there must be more to meaning than truth-conditions.

Are there adequate replies to these challenges?

a. Can translation help? We may say that "renate" is not a translation of "cordate". So we may require that the sentence on the left-hand side of the truth-definition be a translation of the sentence on the right-hand side. But this cannot help. For the notion of translation already presupposes an account of meanings. (Clearly, to understand translation in terms of sameness of truth-conditions is not going to help in this context. For the truth-conditions are the same, and yet we feel that the meaning is different.
b. Davidson: radical interpretation + principle of charity. Try to make maximum sense of the speaker's behaviour. Davidson's thought is that a theory of interpretation will provide the constraints that make the translation correct. So he thinks that he doe not have to rely on a theory of meaning in order to have a theory of translation. Rather, given a correct account of radical interpretation, he can have a theory of correct translation and then a theory of meaning.
   The importance of the **Principe of Charity**: interpret in such a way that you maximise true belief/agreement. Problem in radical interpretation: Suppose you want to radically interpret a Greek who utters "αυτή τη στιγμή βρέχει". Suppose also that it's indeed raining outside. Still, he could mean anything. Or he could utter it even if it was not raining outside. In order to get the right translation, you need to first assume that he has a belief and that this belief is true. In particular, you have to assume that he has the very same belief that you would have under the same circumstances. This is what the Principle of Charity makes possible. You should first attribute to him the belief that it is raining now and then you should assume that his utterance "αυτή τη στιγμή βρέχει" means the same as "it's raining".
   This move would exclude the case of the speaker meaning that snow is white and 2+3=5 when he uttered the sentence "Snow is white". But perhaps meanings are re-introduced by the back-door. What could make say that "p" means p, rather than p and q, if not that "p" Snow is white and 2+3=5 means p?
c. Can we just impose that the sentence on the right-hand side be the same as the sentence on the left-hand side? Not really. First, we still don't understand what the meaning consists in. Second, they are not really the same sentence even if they appear to be (homophonic translation).
d. A more promising answer is to say that an adequate theory of meaning requires not just truth-conditions but also that speakers <u>know</u> the truth-conditions. So, they know that the truth-conditions of "scow is white" is that

snow is white and also they know that this is the strongest (may be the only) condition that makes "snow is white" true. But this answer has unpleasant consequences. One attraction of the truth-conditional theory is that truth-conditions are dissociated from their knowledge. But the present suggestion removes this attraction.

6. What notion of truth is involved in the truth-conditional semantics?

VIII. MEANING AND VERIFICATION-CONDITIONS

➢ Dummett's programme

Dummett's programme is to show that *our* notion of truth—that is, the notion of truth that operates in our language—is an epistemic one. That is, we possess no notion of truth that transcends our capacities to recognise that a particular statement is true (cf. 1982, p.94). His strategy for defending this position has been, in general terms, tied to the ways in which a linguistic community understands and uses a language. Since he thinks that a theory of language use and understanding is a theory of meaning, the strategy is tied to the specification of a semantic theory for a natural language. It is not entirely clear whether this is a descriptive or normative project, but I believe that Dummett understands it as normative. At any rate, what he explicitly says is that a theory of meaning, and hence his own theory of meaning, is a "theoretical representation" of the practice of language users, and it may be such that it proposes a *revision* in that practice (1976, pp.69 & 103).

Dummett's programme has a negative aspect. It concerns the demonstration of the unsuitability of (verification-transcendent) truth-conditional semantics for natural languages. Dummett and his disciples begin with the assumption that a theory of meaning must show what a speaker knows when she understands a language, and what she thereby knows when she understands any given sentence of this language. Then they suggest that truth-conditional semantics fail to show what a speaker knows when she understands a language. Let me sketch their argument—which takes the form of a *reductio*. Suppose that knowing the meaning of a statement consists in knowing its truth-conditions. Knowledge is tied up to an ability to recognise; in particular, knowledge of the truth-conditions of a statement is an ability to recognise that the truth-conditions obtain. Insofar as a language user has a procedure to decide, in a finite time, whether the truth-conditions of a given statement obtain, she knows the meaning of the relevant statement. Yet, natural languages are full of sentences which are not decidable, e.g. statements about the remote past; the future; universally quantified statements where the quantifiers range over infinite domains; statements referring to space-time regions in principle inaccessible to us (1976, p.81). On the face of it, the conclusion is that the knowledge of meaning of these statements cannot possibly consist in knowing their truth-conditions since the latter are beyond recognition (cf. Wright, 1976, p.224).

However, this negative strategy for excluding a truth-conditional theory of meaning in favour for an alternative verification-bound one is not effective for two reasons: *First*, it does not show that statements do *not* have evidence-transcendent truth-conditions. It shows that we cannot always conclusively verify that they obtain. But why is this a problem for a truth-conditional theory of meaning? As Anthony Appiah has argued one can stick to a truth-conditional theory of meaning and at the same time adopt the view that a statement can be said to be true if and only if it is *logically possible* to verify that its truth-conditions obtain. One would then tie truth to the *logical* possibility of verification, but as Appiah has pointed out defending this view, this admission would be compatible with realism. For after all, there is nothing that forces a realist to think that a statement can be true even if there is not even a logically possible difference that its truth would make to our conception of the world (Appiah, 1986). What a realist ought to assert is this: the truth of a statement is logically distinct from *our* epistemic means to verify it. Then, as Appiah stressed: "What is wrong with the argument against realism is not the demand that ascriptions of meaning be verifiable, but the claim that truth-conditional semantics does not satisfy that demand" (op.cit. p.88).

The *second* and most important reason for the actual failure of Dummett's negative strategy comes from challenging the assumption that the Dummettian argument attempts to reduce to absurdity. That is, it comes from challenging the assumption that understanding of the meaning of statements involves knowing the truth-conditions of a statement. In general, as Devitt has argued, understanding a language does not consist in 'knowing that ...', where the truth-conditions of a statement are substituted for the dots. Rather, understanding a language consists in "being able to do things with a language [and] in having thoughts about it" (1984, p.209). More specifically, understanding a language is a "set of grounded skills or abilities" and not propositional knowledge of the truth-conditions of statements (ibid.). So, if realism is not committed to the view that a theory of meaning for a natural language *must* rest on showing how the speakers connect the meaning of a statement with its truth-conditions, then there is nothing in the realist packet for Dummett to refute. Generally, if Dummett's programme is to work it must show muscles. It must not be just a *negative thesis* about truth-conditional semantics. It must provide a *positive alternative* to them. This positive aspect of the programme comes from modelling a semantic theory for natural languages on the intuitionistic semantics for mathematics. Briefly, he general idea of the verification-semantics is this: a competent speaker of L has understood the meaning of a statement S she uses by having *acquired* (internalised, as Putnam would put it) its assertibility conditions; that is the conditions

under which she would be justified to assert S. Similarly, a criterion for understanding of a statement S is that the speaker *manifests*, in her use of S, that she has grasped its assertibility conditions. In other words, what a competent speaker *knows* when she knows a language is the *assertibility conditions* of its statements; i.e. conditions that are not "transcendental" but linked to the knowledge she manifests by her *actual use* of the language. Grasping the assertibility conditions of a statement amounts to recognising whether she is justified to assert that a statement is true. However, assertibility conditions is not just a *sufficient condition* for the truth—realist-style—of the statement in question. Assertibility conditions *replace* truth-conditions; warranted assertibility *replaces* truth—realist-style. When the assertibility conditions of a statement obtain the statement is warrantedly assertible, as opposed to true—realist-style. It follows that verificationist semantics are not bivalent. For it is not the case that every statement is either warrantedly assertible or it is not; or better, it is not the case that if a statement is not warrantedly assertible then its negation is. Yet, as Putnam put it, verificationist semantics leave no semantic "gaps" between statements and the world (1983, p.22): If we cannot conclusively verify a statement, then we abandon the claim that it is either true or false (bivalence); it is neither true nor false. Take, for instance, a statement about the past, say, 'Caesar ate beans before he crossed the Rubicon', that exceeds our capability of verification. A realist could leave a gap: She could say, 'Well, there is a fact of the matter as to whether it is true or false, but there is no way to find it out.' On the contrary, an anti-realist would argue that since we cannot establish the truth of the matter, i.e. we cannot verify the statement, there is no fact of the matter to worry about.

Given this sketch of Dummett's positive programme we may observe that the central issue is whether Dummett has given good reasons to switch to a human-oriented notion of truth. I take it that the only apparent advantage of Dummett's revisionism is that if statements have only verification-conditions, then we are capable of recognising them when they obtain. Then we can also grasp what knowledge of verification-conditions of a statement S consists in: it consists in knowing the circumstances under which it would be warranted to assert that S, the latter assertion being connected with our linguistic practices (cf. Dummett, 1976, p.111).

Is this ability to recognise when the verification-conditions obtain a real advantage of epistemic notions of truth as opposed to non-epistemic ones? I admit that if the realist truth committed us to a largely unverifiable image of the world, or if it were radically and constantly in conflict with warranted assertibility, it would be a problematic notion of truth. But this is by no means the case. What the Dummettians need, generally, is an *independent* argument as to why the alternative conceptions of truth *are* preferable. The fact that epistemic notions of truth give truth 'a human face' is

hardly an independent argument in their favour. For saying that truth must be epistemic in order to be useful to our practices is hardly a non-question begging argument against the realist claim that truth is non-epistemic. Not only Dummettians have not provide such an independent argument , but as we shall now see, their conception of truth runs into severe problems.

An anti-realist would agree that believing, or asserting, something does not, in general, make it true. That is why they talk about *warranted* assertibility, *conclusive* verification and the like. What exactly is the warrant we need to have in order to be able to assert *warrantedly* a statement? The anti-realist conception of truth faces two interrelated problems here. First, it seems to end up with a continuum of epistemic notions of truth, rather than with a single one. Second, it has the awkward consequence that truth is a property that can be *lost*. Let me take them in turn.

Saying that a statement is true iff it can be warrantedly asserted leaves open space for asking the following question: *What is required for a statement to be warrantedly assertible, or for the possibility of its verification?* This question admits many different answers and each way of answering it seems to give different epistemic notions of truth. In order to see that, let me state some of the possible answers:

(a) That *all* of us have *now* conclusive evidence.
(b) That *somebody now* has conclusive evidence.
(c) That *somebody will have* conclusive evidence in the future.
(d) That *some intelligent being has* evidence now.
(e) That *some intelligent being will have* evidence at some future time.
(f) That it is *physically possible* for somebody to find conclusive evidence.
(g) That it is *logically possible* for somebody to find conclusive evidence.

The foregoing list can be augmented by allowing not conclusive, but *prima facie*, evidence; by allowing several communities (e.g., of experts) to be included; even by allowing solipsist solutions (e.g., that it is known by me—now) (cf. Goldman, 1986, p.147). Which exactly is the notion of conclusive evidence that we have to allow in an epistemic conception of truth? And why different communities, or individuals, cannot opt for different 'warranters'? Obviously, different allowances concerning what 'conclusive evidence' stands for generate different epistemic truths and yield different beliefs about the content of the world that can be known as a function of our cognitive capacities. For instance, if current science is taken to give conclusive evidence, then an anti-realist who goes for (b) could say that the statement "three pairs of quarks and leptons were produced after the big bang" is warrantedly assertible. But if current science is not taken to give conclusive evidence, the foregoing statement is not

warrantedly assertible. Moreover, if an anti-realist's answer is (g) then it is not clear at all that a realist could not accept it, without making his notion of truth outrightly epistemic. I do not find a natural way to carve the continuum of epistemic truths at a point. But I think that it is the anti-realist who has the burden of showing how 'conclusive verification' must be understood.

The second problem that epistemic conceptions of truth face is that they make truth a property that can be *lost*. This problem relates to the previous one—that is how to understand conclusive verification. For most of our beliefs, and certainly for almost all of scientific statements, verification can never be conclusive, in the sense the truth of a statement is *entailed* by the evidence at hand. When do we say that the verification of the sentence is conclusive enough so that we can say that this sentence is true, in a time-less way? If we take seriously the anti-realist idea that truth must be tied to verification, evidence and related epistemic notions, there must be some point where a conclusively verified statement entails its own truth. Otherwise, truth would be a property that can be lost, for instance, when fresh evidence disconfirms a statement, or because the conditions of justification, or conclusive verification, change. However, no anti-realist would seriously consider a notion of truth with, as it were, expiry date; for then she would end up with an implausible notion of truth (cf. Smart, 1986, p. 300; Davidson, 1990, p.306). So, the anti-realist needs a notion of epistemic truth that makes it impossible that a statement is true and yet, its truth is going to be questioned.

Dummett has suggested that the anti-realist notion of truth must be such that the notion of conclusive verification involved in it is "of the strongest available kind" (1982, p.95). He rightly thinks that *only if* such a provision is taken into account an epistemic characterisation of truth "rules out the possibility that we shall subsequently be presented with stronger contrary grounds" for a statement that we currently take as warrantedly assertible (ibid.). Hence, he rightly thinks that the anti-realist needs an epistemic notion of truth that makes truth eternal.[1]

However, even if the anti-realist helps herself to an epistemic notion of truth which has it that truth is eternal, she has an important problem to face. It is this: the epistemic warrants that determine truth may be so strong as to rule out as unjustified (and hence untrue) many beliefs that we currently endorse. For if the strength of a warrant is determined by the claim that it rules out the *possibility* that a belief issued by this warrant will be subsequently revised, then quite a number of our current

[1]Recently Wright has suggested a notion of superassertibility as candidate for an epistemic notion of truth (1992, pp.48-61). Superassertibility is “assertibility which would be endurable under any possible improvement to one’s state of information” (op.cit., p.75). Wright correctly suggests that this is the notion of truth that anti-realism needs. But he does not provide an independent argument as to why it is to be preferred over non-epistemic truth.

beliefs would turn out unwarranted. Take, for instance, beliefs based on inductive grounds. How would the anti-realist cope with those beliefs? There are two options available to her. First, she does not regard inductive grounds strong enough so that beliefs acquired by induction are said to be warranted and true. In such a case she shrinks considerably the extent of beliefs that we think they are justifiably true. I do not think that a prudent anti-realist would want to render beliefs acquired by induction unwarranted. Hence, she would follow the second option available to her, viz. to argue that inductive grounds are strong enough so that beliefs acquired by induction are taken to be warranted and true. However, if she does think that induction provides strong grounds for justification, i.e. grounds such that it is *impossible* that a belief is arrived at by induction and yet that this belief may be revised, then, she is in need of an *a priori* argument to the effect that induction is infallible. I do not think that such an argument is available.

Generally, the anti-realist must find a non-epistemic conception of truth that allows that many statements we now believe are true *and* the truth-ascription to them is permanent; i.e. it does not follow the vagaries of our epistemic procedures. In other words, she must show that warranted assertibility is a fixed property of a statement *and* it is possessed by many statements (beliefs) we now entertain. I think that Dummett's attitude to this problem is this: He tries to develop an *a priori* argument to the effect that, at some point, our beliefs cannot help but fit reality. In fact, he states that "it is a priori impossible that most assertions are mistaken" (1982, p.108). The argument is that without taking what another one asserts as true, "we could not have a language" (ibid.). Here, the point is not that most of our assertions *are* mistaken. Rather, the point is how can this be asserted as an *a priori* truth, i.e. a truth beyond any need for empirical justification? Assertions can be even collectively mistaken, (imagine the deceitful demon), and yet we may be able to have a language.

One possible way to justify *a priori* the opinion that beliefs at some point fit reality may come from another *a priori* argument to the effect that the methods we follow to form our beliefs and our subsequent assertions are *a priori* warranted to be infallible. But this is even more difficult to establish. Of course, what the anti-realist would really need to establish is that our methods tend to produce the most conclusive conditions for assertion in the sense that they do not allow the possibility that the grounds on which we assert something now may be weakened later on. The point though is that the methods of inquiry are not God-given and hence they are as contingent and fallible as we are.

It is also worth noting that the anti-realist has another problem to face. If one allows that truth is dependent on our knowledge or recognition of it then one is liable to a relativistic challenge that truth is subjective, community-dependent etc. If an anti-

realist wants to block relativism, that is if she wants to eschew the claim that what statements are warrantedly assertible is community-relative, she is in need of an argument to the effect that there cannot be a diversion in what statements several inquires consider as warrantedly assertible. And if she dismisses the possibility that what is warrantedly assertible is, ultimately, what is true, she has to opt for an *a priori* justification of our methods to the effect that they cannot but lead all of us to the same beliefs, if they are employed correctly.

➢ Putnam's Idealised Justification

Putnam has recently introduced an enhancement to the anti-realist conception of truth (cf. 1981, pp.55-56). He denies that there can be such a thing as conclusive verification, or justification of empirical statements. But he thinks that truth is still epistemic, in the sense that truth is not independent of *all* justification. Instead truth is identified with "idealised justification, as opposed to justification-on-present-evidence" (1983, p.xvii). Since Putnam takes it that ideality is a legitimate notion (think, for instance of 'ideal planes'), he argues that 'ideal epistemic conditions' makes good sense. However, the usefulness of the fiction of 'ideal planes' lies on the fact that the degrees in which ordinary planes approximate ideal ones are specifiable. Hence, by studying 'ideal planes' we can come to know things about ordinary planes. Similarly, if the notion of 'ideal epistemic conditions' is operative, there must be specifiable conditions under which ordinary epistemic situations approximate the ideal ones. I do not have a clue how this can be done, for I do not have a clue what 'ideal epistemic conditions' amount to. Generally, talking of 'ideal conditions' hardly provides a rule as to how and when one is justified in believing that a statement is true. Putnam admits this, but he thinks that it is not a problem. Not only because one must be cautious not to limit in advance what means of verification may become available to human beings; but also because "'truth' (idealised justification) is as vague, interest relative and context sensitive as *we* are" (1983, p.xvii).
Putnam's conception of 'ideal justification" does not improve significantly on Dummett's position. Putnam does not show how truth can be a permanent property of an assertion and yet covary with our epistemic means. Of course, 'ideal conditions for justification' could be understood in a teleological way, that is, as the state of knowledge the human race will be in at the end (or ideal limit) of scientific inquiry. Then, in the face of a 'finished inquiry' truth could be a permanent property of an assertion and yet dependent on our epistemic capacities. But this is not the position that Putnam adopts (cf. 1990, pp.vii-viii). The reason is simple. Putnam wants grounded judgements of warranted assertibility now. He wants his theory of truth to

render true most of our common-sense and scientific assertions. Then, the ideal limit of inquiry is too remote to serve for such a pivot for grounded judgements of warranted assertibility. Putnam's 'ideal justification' is something that can be had now. It refers to *good epistemic circumstances*. Practically, Putnam thinks that a statement is true if it could be justified were epistemic conditions good enough. For instance, there are specifiable conditions under which I can now assert and believe true the statement 'there is a computer in front of me now'. All I would need, for Putnam, is good vision, good lighting conditions, being sober, not having taken hallucinogenic drugs etc.

But there is a certain tension here. If, on the one hand, one thinks of conditions for 'ideal justification' as *available to us now*, then it might be the case that these conditions obtain and yet we acquire a false belief. For instance, it might be the case that all conditions for a justification of the belief that there is a table in front of me obtain, and yet I am a brain-in-a-vat with all table-like experiential inputs being fed to 'me' by a superscientist. Hence, conditions of 'ideal' justification available to us now do not preclude the possibility of error. If, on the other hand, one thinks of conditions for 'ideal justification' as really *ideal*, that is, such that *no human could possibly survey them*, then, as Davidson put it, "they are so ideal as to make no use of the intended connection with human abilities" (1990, p.307). In other words, idealised justification either leaves room for error, and hence it is not at odds with the non-epistemic notion of truth, or, it is so idealised that it is not available to humans. In the latter case also, it would not be at odds with a non-epistemic conception of truth.

➢ A Critique of Minimalism for Truth

Some philosophers—most notably Michael Williams (1986) and Paul Horwich (1990)—*all* that there is in the concept of truth is captured by the disquotational schema

It is true *that* P iff P.

This view has been called 'deflationary' or 'disquotational' or 'minimalist' theory of truth. Its gist is that the truth-predicate does not express any substantial property, like colour or charge. It is there "solely for the sake of a logical need" (Horwich, 1990, p.2). It provides disquotational facilities for sentences (to say that a sentence is true is to assert the sentence) and means for semantic ascent (as in the case of "Every sentence of the form 'p or not p' is true").

Deflationists argue that *all* the facts involving truth can be explained on the basis of the minimal theory (cf. Horwich, 1990, p.7). Of course, the disquotational schema is not enough to account for all the uses of the predicate 'is true', as it is clear from the case "All consequences of true sentences are true". But then again the deflationist would argue that the disquotational schema is *near enough* wholly explanatory of the uses of the truth-predicate in a language. According to Horwich the minimalist theory of truth has just a superficial resemblance to the Tarski theory of truth (1990, pp.42 & 120). This is correct since Horwich eschews the Tarskian explanation of truth in terms of satisfaction. Instead his own theory of truth consists of (an infinite number of) axioms of the form 'It is true *that* P iff P' where P is a variable ranging over propositions (expressed by sentences of a language L) (op.cit. pp.5-6 & 21).
What I want to argue here is that the minimalist conception of truth, at least as formulated by Horwich, faces some problems that cast doubt on its adequacy and give us reasons to believe that a substantial notion of truth is needed. For a start, Horwich's theory faces problems when the object language is not contained in the meta-language. If the two languages are different, issues of translation from one language to another must be taken into account and as Donald Davidson put it, the concept of translation "is far more obscure than that of truth" (1990, p.296). Let us see what exactly the problem is. Horwich suggests that the following schema exhausts the nature of truth for utterances

(D-tr) **u** is true iff p,

where '**u**' is replaced by a singular term referring to an utterance and 'p' is replaced by a sentence of our language that would be a *translation* of that utterance (1990, p.106). Translation, for Horwich, is a procedure that is determined by "sameness in overall use" and is such that is characterised in terms that do not involve truth (op.cit., p.99). This latter point is crucial, for if translation is to involve truth (i.e. sameness in truth-conditions) the whole process of minimising truth is jeopardised. This observation was made vividly by Harty Field: He pointed out that if translation was dependent on truth "the notion of truth would be needed to determine which are the appropriate instances of the schema [D-tr] that is supposedly explaining what truth is" (1992, p.324). Then, the issue is this: is "sameness in overall use", as distinct from having the same truth-conditions, adequate for translation from an utterance in a foreign language to ours? In particular, the issue is this: Does Horwich explain "sameness in use" in a way that avoids mention of sameness in reference? Does he appeal to sameness-of reference devices other than sameness in the causal networks that connect the foreign utterance **u** and its referent with its translation in our language and

its referent? (cf. Field, ibid.) Unfortunately, Horwich does not explain what "sameness in overall use" amounts to. Prior to such an explanation though, his claim that translation eschews citation of truth-conditions cannot be assessed. It remains a promissory note. Moreover, Horwich admits that "claims of inter-translatability based on similarity of use are not completely determinate" (1990, p.101). I think that the point here must not be one of ignorance: either two words *are* inter-translatable or they are *not*, but overall use cannot decide it. The point is that *use* may radically underdetermine translation in the same way as use of rigid rods underdetermines whether a 'geodesic' should be taken as straight line in the Euclidean sense or as line with positive curvature. Contrariwise, translations in terms of truth-conditions and sameness in reference give us a uniform way of translating; although, at a given point we may be ignorant as to whether we possess the correct translation. But translations in terms of truth-conditions and sameness in reference pull away from the minimalist conception of truth and require a substantial notion of truth, as Field rightly observed. Hence, the minimalist theory of truth seems inadequate to explain the nature of truth for utterances when translation is involved, whereas a substantial theory of truth would do the job.

The minimalist theory of truth faces other problems. For instance, Field has argued that minimalism cannot account for the role that truth-conditions play in explaining the various agents' success and failures in getting about the world (1992, p.329). A substantial explanation of one's successes would be that *agents tend to be more successful when the truth-conditions of their beliefs are realised*. What would the minimalist alternative be? Horwich could appeal to some use-related aspect of a belief state as explanatorily relevant to its success, e.g. its verification-conditions. I do not want to argue that an explanation such as *agents tend to be more successful when the verification-conditions of their beliefs are realised* cannot be framed. Rather the point is that such an generalisation would be in need of substantiation. It is not that obvious that there is a correlation between verification-conditions and success. Much depends on how exactly one understands verification and the realisation of verification-conditions. Moreover, one can argue that the verification-conditions of a belief may be realised and still action based on these conditions may not be successful. Imagine, for instance, beliefs concerning unobservable entities. For example, let us imagine a person who believes that humanity must not rest until the last small pox virus has disappeared. In that case, the verification-conditions of the relevant belief may obtain and yet, clearly, success is warranted only when the truth-conditions of this belief are realised. At any rate, talking about verification-conditions involves a substantial, yet epistemic, understanding of truth. Hence it would not really be to Horwich's avail. So, minimalism cannot be the whole truth and nothing but the truth about truth.

➢ The Model-Theoretic Argument Against Non-Epistemic Truth

Putnam's model-theoretic argument seems to offer independent grounds for denying the non-epistemic character of truth. In fact Putnam has argued that non-epistemic truth becomes *incoherent* and *unintelligible* (1978b, 124 &126). Putnam envisages a thought experiment. He takes an ideal theory T of the world; a theory that satisfies all operational and theoretical constraints; that possesses any property that we can imagine or please except objective truth—which is left open. He takes it that for a realist—what Putnam calls 'metaphysical realist'—T might (in reality) be false. His argument then is a *reductio*: if we assume that 'ideal T' might still be false we end up with absurdity. Let me sketch the argument.

Suppose that T says that there are infinitely many things in the world. T is consistent (by hypothesis) and it has only infinite models. By the Löwenheim-Skolem theorem, T has a model of every infinite cardinality (greater than or equal to the cardinality of extra-logical symbols of the language of T). Now, pick a model M of T, having the same cardinality as the WORLD. Devise an one-to-one mapping **m** of the individuals of M onto the pieces of the WORLD and map the relations between the individuals of M directly into the WORLD. These mappings generate a satisfaction relation (in Tarski's sense)—call it SAT*—between (sets of) pieces of the WORLD and terms and predicates of the language of 'ideal T' such that T comes out true of the WORLD. (That is, the WORLD is isomorphic to the model M in which T is true.) The ideal theory has been shown to be true of the WORLD. Then how can we claim that the ideal theory might really be false? (Putnam, 1978b, 126; cf. also 1981; 1983) This would be just absurd.

Nonetheless, the realist could argue that the satisfaction relation SAT* that was devised as above is not the *intended* correspondence between the theory and the WORLD. Putnam reminds the reader that 'ideal T' satisfies—by hypothesis—*all* operational and theoretical constraints: it squares up with all observations; it is simple, unified, and the like. Above all, SAT* satisfies an important constraint on reference: it makes T come out true. Then, he offers the realist a challenge: How SAT* could be not intended? What further constraints could show that SAT* is not intended; or, equivalently, that some *other* interpretation is the intended one? Putnam's own conclusion is that nothing in the world could show that SAT* is *not* the intended interpretation, and that some else—call it SAT—is. He also adds that since *any* a model of T which is isomorphic to the WORLD would, by construction, satisfy all operational and theoretical constraints, there would be no sense in which only one of those *is* the intended one. Instead, since operational and theoretical constraints are the

only means to single out the intended model, all models of T that satisfy them would be intended, and the ideal theory T would be true in all intended models! (1983, 13) In other words, not only can we not find out a unique truth-preserving interpretation of a language, but

> there are always infinitely many different interpretations of the predicates of a language which assign the 'correct' truth-values to the sentences in all possible worlds, no matter how these 'correct' truth-values are singled out (1981, 35).

Putnam's argument seems powerful. For although a realist may immediately say that the argument does *not* prove that there is not a unique intended interpretation of a language, (there may be even if we shall never come to specify it), Putnam's challenge is that the very notion of a unique interpretation fixed by the world—implicit in a non-epistemic theory of truth—makes no good sense. Moreover, Putnam quickly anticipated an objection that a realist would have. A realist could object that a causal theory of denotation would show that, and explain why, a particular referential scheme for a language L—call it the intended interpretation—is picked out, thereby refuting Putnam's argument. However, Putnam argued, a causal theory of reference would be of no help to the realist. For the model-theoretic argument can be extended to words like 'cause': 'cause' can be reinterpreted no less than other words; in each model M, reference_M will be defined in terms of cause_M. Then, Putnam said,

> unless the word 'causes' is already glued to one definite relation with metaphysical glue, this does not fix a determinate extension for 'refers' at all (1983, 18).

'Reference' and 'cause' do not enjoy the privilege of an unique intended interpretation; they are also referentially indeterminate.

Is Putnam's argument, after all, such a powerful and all-conquering one? As Clark Glymour (1982, 177) has nicely put it, Putnam's argument seems to be "a kind of endless dialogue": whenever one says something about what singles out a referential scheme, Putnam says it is insufficient to do the job, for what one says "adds more theory" which may be re-interpreted in countless ways, and hence it is itself referentially indeterminate. But, fortunately for the realist, Putnam's argument is instructive—as any good argument must be—but faulty. Despite Putnam's apparent neutralisation of the role of causal considerations in fixing the intended interpretation, the real damaging response to the model-theoretic argument comes from such

considerations.[2] Let me show how the realist answer goes, by illustrating first how the supposed "endless dialogue" works.

The central technical idea of the model-theoretic argument is that of *isomorphism*. In particular, the argument assumes that the WORLD consists of an infinite number of individuals which can be mapped upon the individuals of the universe of the model M of the ideal theory T —with the same cardinality as the WORLD—in such a way that T becomes true of the WORLD. Under this assumption, the argument goes through nicely. But, as G H Merrill (1980) observed, it is questionable whether realism conceives of the world merely as *a set of individuals*, i.e., as a model-theoretic *universe of discourse*. The world, a realist would say, is a *structured entity*. Its individuals stand in specific relations to one another, or to subsets of individuals. In other words, the model-theoretic description of the world must *not* be just a universe of discourse, but rather a *structured universe of discourse*. In particular, a realist would assert two things: (1) the constituents of this universe (individuals and properties/relations) are independent of any particular representation we have of it (op.cit. 72); (2) whereas Putnam's assumption is that the language precedes the universe (domain of discourse) and 'structures' it, the realist position is that the universe is *already* structured, *independently* of the language. Then, a realist could argue that the model-theoretic argument fails. For, an interpretation of the language, i.e., a referential scheme, either matches the language to the existing structured universe or it does not. If it does not, then there is a clear-cut case in which even an ideal theory *might* be false. In particular, if the WORLD is a structured domain, then in order for Putnam to have his model-theoretic argument, he would have to show that the mappings from a model M of T onto the WORLD are structure-preserving. Yet, it simply is *not* always possible to produce structure-preserving isomorphisms (op.cit. 74).

So, the realist seems able to block the model-theoretic argument. However—and here is where Putnam's 'endless dialogue' comes into play—the realist way-out can be itself, as it were, 're-interpreted'. As Merrill and Gregory Currie (1982) observed, the realist claim that it is *not* always possible to produce structure-preserving isomorphisms between a model M (of the cardinality of the WORLD) and the WORLD rests on the following prior assumption: some subsets of the power set of the

[2] An interesting way to attack the model-theoretic argument is via attacking the very notion of an ideal theory T, i.e., the central premiss of Putnam's thought experiment (cf. Koethe, 1979; Tuomela, 1979). The thrust of the objection is that Putnam cannot help himself to the notion of ideality. For his premiss is that T is an ideal theory, "*by our lights*" (1978b, 125—emphasis added). Then, for all *we* know, no theory T can be really ideal in that it cannot be *revised*; i.e., in that there can be no theory T' after T such that T' is inconsistent with T, or simpler than T, or more unifying than T etc. For all *we* know, any theory that we might come to believe by *our* lights, is going to be revisable.

universe of discourse of the WORLD are *privileged*, in the sense that they specify extensions of the properties of individuals of the WORLD. These subsets provide, in essence, the structure of the universe of discourse. If this is so, then there is no warrant that the WORLD can always be made isomorphic to another model M with the same cardinality as the WORLD. For simply, it is not always possible to construct a model such that it is the mirror-image of all and only the properties, relations and individuals of the WORLD. On the contrary, if the WORLD has no structure, or if it is "maximally structured"—i.e., if it contains *all* subsets of the power set of its universe of discourse—then it will be sufficiently rich to be mapped, in a structure-preserving way, onto *any* model with the same cardinality. In other words, the requirement of a structured domain presupposes some distinction between *properties* (e.g., blue, green etc.) and *pseudo-properties* (e.g., grue, bleen) such that the former belong to the structure of the WORLD and single out subsets of its universe of discourse as their extensions, whereas the latter do not belong to the structure of the WORLD and their extension do not single out properties of the world.

The relevant issue here is what exactly distinguishes between properties and pseudo-properties, e.g., blue and grue. Currie suggested that any proposal as to what distinguishes between properties and pseudo-properties amounts to a provision of a constraint, call it C, upon the semantics of T (1982, 264). But then, he said, we are back in square one: we add *more theory* to our theory of the world. The conjunction of T and the semantic theory of C—call it CT—has a model, if both T and C are consistent. Hence, Putnam's model-theoretic argument applies to CT, "in such a way as to establish that T is true of the real world *and* that it is true in a faithful interpretation" (ibid.). So, the realist way-out is neutralised by Putnam's 'endless dialogue'. Or is it not?

Lewis (1984) shows that Putnam's "never-ending dialogue" can stop. His point is that when one suggests an extra constraint—call it *C*-constraint—in order to fix the intended referential scheme, what one suggests is that *in order for an interpretation to be intended, it must conform to C*. Then, the real issue is not whether the *theory* of *C* will come out true under unintended interpretations. Rather, it is what exactly *C* is and how it operates. In the light of this, Lewis's suggestion is that the appeal to causal considerations in fixing the intended referential scheme is not just adding more theory but offering *constraints* to which an interpretation must conform in order to be intended.

Lewis's point explains precisely what the causal theorists, such as Devitt (1983), (1984), (1984a) and Glymour (1982), mean when they say that "causal glue" fixes the intended interpretation. They introduce a causal constraint in reference-fixing to which an interpretation must conform. Then the issue is not that the theory of this

constraint may be re-interpreted, but rather why and on what grounds we choose this constraint rather than another. Putnam's objection that the reference of the two-place predicate 'causally related to' turns out itself indeterminate is, then, easily overruled. To put it crudely, Putnam's objection conflates use and mention (Glymour, 1982, 177). Reference is not fixed by theories that contain the predicate 'causally related to'. Rather, reference is fixed by *causal relations*. The predicate 'causally related to' is just another predicate of our theory of the world (which includes the linguistic formulation of the causal theory of reference); it gets its reference in the way that any other term and predicate gets it. More generally, a causal theory of reference will be expressed, trivially, in words. But causal relations do no require linguistic representations in order to operate. Putnam's argument has challenged the realist to show how the intended interpretation of a theory T is fixed. The realist can do that; that is, she can *explain* how the intended interpretation is fixed by causal considerations. Her explanation is bound to have a linguistic formulation. But this inevitable feature of our theorising does not entail that the thing that does the explaining, i.e., causal connections, is nothing but words.

Lewis also shows how Merrill's view of the WORLD as a *structured domain* can be defended against Currie's argument. He suggested a *constraint* on the interpretation of a language, viz. that referents must be *eligible*. He conceded that one can posit putative things and classes of things at will. However, among the countless things and classes that can be posited,

> most are miscellaneous, gerrymandered, ill-demarcated. Only an elite minority are carved at the joints, so that their boundaries are established by objective sameness and difference in nature. Only these elite things and classes are eligible to serve as referents (1984, 227).

In other words, Lewis idea is that from the countless things that can be posited, only a minority is such that carves the world at its joints, in the sense of picking out individuals and properties which belong to the objective structure of the world. Lewis's eligibility-constraint is *prior* to the interpretation of a language and such that an interpretation must satisfy in order to be intended. Lewis does not suggest an eligibility-theory which is open to re-interpretations. He offers a constraint and suggests that we have to turn to physics and its 'inegalitarianism' in order to find the elite things and classes of things that constitute the joints of the world. Then he argues that an interpretation is unintended—and it would be disqualified—if it employs gerrymandered referents; i.e., putative referents that do not belong to the objective structure of the world. It follows that if "we limit ourselves to the eligible

interpretations, the ones that respect the objective joints in nature, there is no longer any guarantee that (almost) any world can satisfy (almost) any theory" (1984, 227).