Abstraction and Semantic Explanation

William Starr

Cornell University
Sage School of Philosophy
will.starr@cornell.edu
http://williamstarr.net

April 5, 2012

What I'm Talking About
'Semantics'

'Semantics'

"The word 'semantics' is used here in a narrower sense than usual. We shall understand by semantics the totality of considerations concerning those concepts which, roughly speaking, express certain connexions between the expressions of a language and the objects and states of affairs referred to by these expressions." (Tarski 1936b: 401)

"Alfred Tarski" ⟶ 

Semantics
And Semantic Theory

Semantics (Tarski 1936a)

A semantics for a symbolic system says what thing in the world each symbol is *about* (objects, states of affairs, etc.)

- Familiar, intuitive idea: words/thoughts are used to talk/think about things
- It's nice to vindicate our common sense opinions...
- But why think of semantics as a predictive theory?
- Maybe: there are regularities in what our words/ideas are, intuitively, about
  - E.g. "Mars is red" is about a state of affairs involving the thing called "Mars" and the property called "red"

A Common View
About Semantic Theory

A Common View of Semantic Theory

1. A semantic theory captures regularities in what symbols are, intuitively, about
2. Regularities principally exhibited by complex symbols
   - Explanation: they inherit their 'aboutness' from their constituents (+syntactic structure)
3. Supplement: naturalist account of 'aboutness' for atomic symbols (e.g. conventional, informational)
   - Appendix: theory of how atomic symbols come to stand in this relation (metasemantics)
4. Application: a semantic theory captures (part of) an agent's competence with a symbolic system

Three Objections to the Common View

Objection 1: Deflationism (E.g. Field 1994, Horwich 1990, 1998)

- In capturing these regularities a semantic theory ends up saying things like:
  - The referent of "Tarski" is Tarski
- We can maintain the intuition that these claims are true without construing reference as a relation between things and the external world
- **Deflationism**: People just treat "the referent of "Tarski"" and "Tarski" as synonymous
- So our basic intuition just comes to: Tarski is Tarski
  - Talk of 'reference' is eliminable
  - Intuitions get paid lip service, w/o genuine symbol-world relation

Three Objections to the Common View

Objection 2: Chomskian Skepticism (Chomsky 1995)

- On the common view, semantics amounts to predicting people's opinions about reference, truth, etc.
- Why think this pertains much to people's distinctively <u>linguistic</u> competence?
  - Their capacity to generate and interpret linguistic structures?
- Why not cognitive psychology and ethnoscience?
- Further, people talk about all kind of 'things' which aren't real objects in the world
  - "We need to find a location for al-Quds"
- So intuitive 'reference' and the theoretical 'reference' of semantic theory are clearly distinct

Three Objections to the Common View

Objection 3: Inappropriate References (Starr ?)

- Suppose our intuitive judgements of reference and truth *are* reflections of our linguistic competence
- It seems uncontroversial that perfectly competent speakers can fail to know what their words are about
  - It was a *discovery* that "water" and "H_2O" pick out the same substance
- Regularities emerge with complex structures: *p and q*
  - But it is immanently *unintuitive* to think of words like "and" as referring to something
- Dilemma:
  - The regularities aren't explained by 'aboutness'
  - Or limit the theory's scope to atomic expressions and rob it of its motivating regularites

Outline of the Talk

1. Argue: the Common View does not match the semantic explanations given in logic and cogsci
2. Articulate the view that does based on *information* and rebuts Objection 1
3. Note this view doesn't yield a plausible account of 'symbolic competence' (doesn't meet Objection 3)
4. Make some observations about abstraction and information processing in semantic explanations
5. Use these to inspire a toy semantic theory which does meet Objection 3
6. Address Objection 2

The Rest of the Talk

Background on Tarski's Project
Proof and Logical Consequence

- Tarski was concerned with the study of artificial languages like first-order logic
  - Precise, simple syntax:
    - Atomics: e.g. $\mathsf{Red(mars)}$
    - Compounds: $\mathsf{A} \wedge \mathsf{B}, \neg\mathsf{A}$
- Purpose: representing deductive arguments (proofs)
  - A sequence of sentences starting from premises and ending in a conclusion
- **Goal**: a precise account of which claims follow from others, of logical consequence

Tarski's Project
A Formal Definition of Logical Consequence

Tarski's Goal

Definition of consequence that correctly identifies which formulas follow from which (*notation*: $\mathsf{P}_1, \ldots, \mathsf{P}_\mathsf{n} \vDash \mathsf{C}$)

- Formal, 'internal', account:
  - Good arguments are sequences which conform to legitimate formal rules of inference
    - E.g from $\mathsf{A} \wedge \mathsf{B}$ you can deduce B
  - $\vDash$ relation reduced to relations between symbols
  - Example: $\mathsf{A} \wedge \mathsf{B} \vDash \neg\neg\mathsf{B}$ reduces to:
    - Deducibility of B from $\mathsf{A} \wedge \mathsf{B}$
    - And deducibility of $\neg\neg\mathsf{B}$ from B
  - In short: reduce consequence to formal proof

Tarski's Project
A Formal Definition of Logical Consequence: Problem

Tarski's Goal

Definition of consequence that correctly identifies which formulas follow from which (*notation*: $\mathsf{P}_1, \ldots, \mathsf{P}_\mathsf{n} \vDash \mathsf{C}$)

- Problem: 'incompleteness' (Gödel 1931; Tarski 1939)
  - If formal language is as expressive as arithmetic, there are consequences for which no proof exists
  - That is: $\mathsf{P}_1, \ldots, \mathsf{P}_\mathsf{n} \vDash \mathsf{C}$ but C is not formally provable from $\mathsf{P}_1, \ldots, \mathsf{P}_\mathsf{n}$
- The relation of interest doesn't reduce to system-internal relations (Tarski 1936a)
- So: Deflationism objection is misguided (Shapiro 1998; Ketland 1999)

Tarski's Project

The Genesis of Semantics

- Tarski's semantic hypothesis:
  - Instead reduce ⊨ to relations holding between the symbolic system and 'the world'
- Old idea (Frege, Wittgenstein, Carnap):
  - Logical consequence is about truth preservation: it's impossible for premises to be true if conclusion is
  - Impossible: no way of 'being true'
- New idea:
  - P(n) is true if and only if the referent of 'n' is among the things to which 'P' applies
  - Ways of being true: a way of
    1. Referring to objects with names and referring to properties with predicates
    2. Objects having properties

Tarski's Semantic Account of Consequence

A Convenient Formal Restatement

- A *world* w:
  - What things are there? $D_w = \{d_0, \ldots, d_n, \ldots\}$
  - What are those things like? Groupings of 'like objects': $G_w = \{\{d_i, \ldots, d_j\}, \ldots, \{d_k, \ldots, d_l\}\}$
- A reference relation r between symbols and the world:
  - r pairs names w/objects; predicates w/a set of objects in each world (a property)

Example $w_1 : D_{w_1} = \{a, b\}, P_{w_1} = \{\{a\}, \{b\}\};$
$w_2 : D_{w_2} = \{a, b, c\}, P_{w_2} = \{\{b, c\}, \{a\}, \{b\}, \{c\}\}$

1. $r(\mathsf{tarski}) \longrightarrow a$
2. $r(\mathsf{Ran}) \longrightarrow \begin{cases} w_1 \longrightarrow \{a\} \\ w_2 \longrightarrow \{b, c\} \end{cases}$

Tarski's Semantic Account of Consequence

A Convenient Formal Restatement

Example $w_1 : D_{w_1} = \{a, b\}, P_{w_1} = \{\{a\}, \{b\}\};$
$w_2 : D_{w_2} = \{a, b, c\}, P_{w_2} = \{\{b, c\}, \{a\}, \{b\}, \{c\}\}$

1. $r(\mathsf{tarski}) \longrightarrow a$
2. $r(\mathsf{Ran}) \longrightarrow \begin{cases} w_1 \longrightarrow \{a\} \\ w_2 \longrightarrow \{b, c\} \end{cases}$

- According to r, Ran(tarski) is true in w iff
  - $r(\mathsf{tarski}) = t$, $r(\mathsf{Ran})(w) = R$ and $r \in R$
- According to r, Ran(tarski) is:
  - True in w_1
  - False in w_2

Tarskian Semantic Consequence

Payoff

Tarskian Consequence

$\mathsf{P}_1, \ldots, \mathsf{P}_\mathsf{n} \vDash \mathsf{C}$ iff every r and w that make $\mathsf{P}_1, \ldots, \mathsf{P}_\mathsf{n}$ all true also makes C true

- I've said nothing about the truth complex sentences...
- Suppose I don't: what regularities in truth-conditions (patterns of consequence) would this theory capture?
  - NONE!
- Those patterns emerge after adding clauses like this:
  - Only way to make $\mathsf{A} \wedge \mathsf{B}$ true is to make A and B true
- But these clauses don't really belong to the semantic theory and don't mention reference at all

Tarskian Semantic Consequence
Objection 3

- I've said nothing about the truth complex sentences...
- Suppose I don't: what regularities in truth-conditions (patterns of consequence) would this theory capture?
  - NONE!
- Those patterns emerge after adding clauses like this:
  - Only way to make $\mathsf{A} \wedge \mathsf{B}$ true is to make A and B true
- But these clauses don't really belong to the semantic theory and don't mention reference at all

Tarskian Semantics
One Response to Objection 3: Thinking Like Frege

- Extend the notion of 'reference' to sentences:
  - $[\![\mathsf{P}(\mathsf{n})]\!] = \{\langle w, r\rangle \mid \text{given } r, \mathsf{P}(\mathsf{n}) \text{ is true in } w\}$
  - An atomic sentence picks out all of the reference/world assignments that make it true
- Gives a reasonable picture of sentence meaning:
  - Information: signals distinguish states the world could be in (Shannon & Weaver 1963; Dretske 1981)
- Now, stretch the notion of reference even further: $\wedge$ refers to the function of set intersection
  - $[\![\mathsf{A} \wedge \mathsf{B}]\!] = [\![\mathsf{A}]\!][\![\wedge]\!][\![\mathsf{B}]\!]$ and $[\![\wedge]\!] = \cap$
  - $[\![\mathsf{A} \wedge \mathsf{B}]\!] = [\![\mathsf{A}]\!] \cap [\![\mathsf{B}]\!]$
- Now reference is playing a role!

Tarskian Semantics
Objection 3 Stands

- The response seems cheesy: the symbol/world relation has been stretched beyond recognition
- Information is a sensible and central concept, but kind of thing we've taken to be the referent's of $\wedge$... less so.
- Further, we still seem to be equating symbolic competence w/how a symbol is connected to the world
- And that seems immanently inappropriate

Marr's *Vision*
Outline

- Three leveled approach to explaining vision:
  1. Computational (mathematical) theory
  2. Representational/algorithmic implementation
  3. Hardware implementation
- Three-Leveled Methodology:
  1. What is the mathematical problem is being solved?
  2. How is it solved?
     - What representation/algorithm does it use?
  3. How could neurons implement that solution?
- How does Marr motivate the computational level?

The Representation Level
Are Representations and Neurons Enough?

- Neurophysiological and representational theories of vision are both clearly necessary
- Why not stop there?
- Computer vision researchers studied vision in the 'white blocks world'
- They devised representations & algorithms for deriving representations of 3D objects from line drawings
- But methods didn't scale up
- What was needed: more abstract characterization of what vision involved and a representational approach that captured all aspects of that
- In short: a goal for the representational implementation

A Complete Theory
Needs To Say What is Being Done and How

To Clarify (Marr 1982: 5)

"...[M]ost analogies between brains and computers are too superficial to be useful. Think, for example, of the international network of airline reservation computers, which performs the task of assigning flights for millions of passengers all over the world. To understand this system, it is not enough to know how a modern computer works. One also has to understand a little about what the aircraft are and what they do; about geography, time zones, fares, exchange rates, and connections; and something about politics, diets, and the various other aspects of human nature that happen to relevant to this particular task."

A Complete Theory
Needs To Say What is Being Done and How

The Point (Marr 1982: 5)

"Thus, the critical point is that understanding computers is different from understanding computations. To understand a computer, one has to study that computer. To understand an information-processing task, one has to study that information-processing task. To understand fully a particular machine carrying out a particular information-processing task, one has to do both things. Neither alone will suffice."

The Computational Level
Example 1: Color

- Color is a perceptual approximation of reflectance
- A goal of color vision:
  - Separate the effects of reflectance changes from the vagaries of the prevailing illumination
- A solution (Horn 1974): Reflectance changes are sharp, illumination changes gradual
  - Filter out slow changes, rest is due to reflectance
  - There is an equation that factors out the slow changes from the total intensity changes
- This goal and solution are stated in a way that are independent of how color vision is made to work at the representational and implementation levels

The Computational Level
The Gist

The Examples

"Gone is the restriction to a special visual miniworld; gone is any explanation *in terms of* neurons — except as a way of implementing a method." (Marr 1982: 18)

The Message (Marr 1982: 19)

"The message was plain. There must exist an additional level of understanding at which the character of the information processing tasks carried out during perception are analyzed and understood in a way that is independent of the particular mechanisms and structures that implement them in our heads... Such analysis does not usurp an understanding at the other levels — of neurons or of computer programs — but it is a necessary complement to them..."

The Computational Theory
Is a Mathematical Theory

- What are the components of the theory at this level?
- Color vision: equation relating two quantities
- Shape vision: equation factoring one quantity into two and disregarding one
- "I have argued that from a computational point of view [the retina] signals $\nabla^2 G * V$ (the X channels) and its time derivative $\partial/\partial t(\nabla^2 G * V)$ (the Y channels). From a computational point of view, this is a precise characterization of what the retina does." (Marr p.337)

The Point The computational level of the theory consists of some mathematical equations

The Computational Level
Representation, Finally

- Showing that the equations are appropriate:
  - Showing that they in fact track properties of the world important to the task of vision
  - Properties like shape, depth and color
- That is, the equations must contain variables which covary with features of the environment
  - And those features must be used to see
  - Used to know what is where by looking
- Without this step the process characterized by the equation is not describing vision

The Levels Methodology Applied
The Cash Register (pp.22-4)

Computational Level: What & Why? (Step 1a)

- A cash register does arithmetic
- So, we start w/an abstract theory of what that is
- Addition is a mapping ('+') of two numbers to one:
  - $(3 + 4) \mapsto 7$
- It's commutative: $3 + 4 = 4 + 3$
- It's associative: $(3 + 4) + 5 = 3 + (4 + 5)$
- There's a zero: for any n, $0 + n = n$
- Each n has a unique inverse $-n$ & they add to 0: $n + (-n) = 0$

Important: all of this is true regardless of the representational system (Arabic, Binary, Roman)

The Levels Methodology Applied

The Cash Register (pp.22-4)

Computational Level: What & Why? (Step 1b)

- A cash register does arithmetic, but why?

  Zero If buy nothing, cost nothing, and buying nothing and something should cost the same as buying just something.

  Commutivity The order in which any two goods are rung up does not matter

  Associativity Any way of arrange goods in piles and paying separately won't change total cost

  Inverses If you buy something then get a refund, the total cost should be 0

The Levels Methodology Applied

The Cash Register (pp.22-4)

Marr on the Computational Level (p.23)

"This whole argument is what I call the *computational theory* of the cash register. Its important features are (1) that it contains separate arguments about what is computed and why and (2) that the resulting operation is defined uniquely by the constraints it has to satisfy."

A Suggestive Analogy

The Cash Machine and Natural Language...

- Think about the difference between entering '11.99' and '+' on the cash register
  - The first has the job of indicating external value
  - The second processes information received
- It seems essential to say that the internal states of the machine carry information about value
  - But that the machine's competence resides in it's capacity to processes this information
- Its competence with '11.99' involves using it to gain information about value
  - Its competence with '+' resides in processing that information

Starting Point

Information and Tarski

- Think of information as some physical state which covaries with another
  - Information thereby distinguishes some states of the world from others
- Sets of w/r pairs, call them i, abstractly characterize this:
  - The pairs in the set vs. the pairs excluded
- So the goal is to specify the meanings of sentences in terms of how they can be used to process information
- Mathematically: associate each sentence with a function from one i to another i'1

Dynamic Semantics

Natural Language as a Programming Language

Dynamic Semantics for FOL

1. $i + \mathsf{P}(\mathsf{n}) = \{\langle w, r\rangle \in i \mid r(\mathsf{n}) \in r(\mathsf{P})(w)\}$
   - Eliminate the possibility that the referent of n doesn't have the property picked out by P
2. $i + \mathsf{A} \wedge \mathsf{B} = (i + \mathsf{A}) + \mathsf{B}$
   - Update w/information carried by A, then update w/information carried by B
3. $i + \neg\mathsf{A} = i - (i + \mathsf{A})$
   - Eliminate the information that A

Formal Inspirations: Pratt (1976); Heim (1982); Veltman (1996)

Dynamic Semantics

And Objection 3

- You can be competent with n w/o knowing what it actually refers to
  - One way: a state of referential uncertainty $\{\langle w, r'\rangle, \langle w, r\rangle\}$ where $r(\mathsf{n}) = a \neq b = r'(\mathsf{n})$
- Just as error can lead to the exclusion of the actual world, it can leave you with a nonactual reference relation
  - $r(\mathsf{water}) \neq r(\mathsf{H_20})$
- This isn't linguistic incompetence, it's just false belief

Dynamic Semantics

Chomskian Skepticism

- I've been calling i information, which suggests veridicality
  - i does not exclude the actual world
- But it generally isn't
- For purposes of modeling communication, it should be thought of what we are mutually taking for granted for the purposes of the conversation
  - Mutually: I'm taking it for granted, you are too, I'm taking it for granted that you are taking it for granted, etc. (Stalnaker 1970; Clark 1996)
- We can be fantastical, pretend, act, let each other speak loosely, etc.

Dynamic Semantics

Chomskian Skepticism

- In these cases, we may be able to convey some real information about some topic, while saying something false, non-sensical about others
- We may presuppose odd ontology to meet our ends and so on
- It is only in scientific contexts that we aspire to using language and concepts in a way that is ontologically committing
- So, we can talk about the strangest things in some contexts, and know what we mean, even if none of these things exist

The Semantic Thesis

Of The Talk

1. Symbolic systems do have semantics in Tarski's sense
   - Some symbols carry information (by convention or natural law)
2. But this alone does not capture the competence of those who use that symbolic system
   - Use = information processing
   - Competence = constraints on information processing
3. The two styles of 'semantics' can be embraced simultaneously when are explanations are formulated at a 'wide' level of abstraction
   - But this is only part of the story
   - We also need a representational and a physical explanation

Acknowledgments

Thank You!

Slides at http://williamstarr.net/research

References I

Chomsky, N (1995). 'Language and Nature.' *Mind*, **104(413)**: 1–61.

Clark, HH (1996). *Using Language*. Cambridge, England: Cambridge University Press.

Dretske, FI (1981). *Knowledge and the Flow of Information*. Cambridge, Massachusetts: The MIT Press.

Field, H (1994). 'Deflationist Views of Meaning and Content.' *Mind*, **103(411)**: 249–285.

Gödel, K (1931). 'Über formal unentscheidbare Satze der Principia Mathematica und verwandter Systeme I.' *Monatschefte fur Mathematik und Physik*, **38**: 173–198.

Heim, I (1982). *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, Linguistics Department, University of Massachusetts, Amherst, Massachusetts.

Horwich, P (1990). *Truth*. Oxford: Basil Blackwell Publishers.

Horwich, P (1998). *Meaning*. Oxford: Oxford University Press.

Ketland, J (1999). 'Deflationism and Tarski's Paradise.' *Mind*, **108(429)**: 69–94.

References II

Pratt, VR (1976). 'Semantical Considerations on Floyd-Hoare Logic.' In *Proceedings of the 17th IEEE Symposium on Foundations of Computer Science*, 109–121.

Shannon, CE & Weaver, W (1963). *The Mathematical Theory of Communication*. Urbana, Illinois: University of Illinois Press.

Shapiro, S (1998). 'Proof and Truth: through Thick and Thin.' *Journal of Philosophy*, **95(10)**: 493–521.

Stalnaker, RC (1970). 'Pragmatics.' *Synthese*, **22**: 272–289. 10.1007/BF00413603, URL `http://dx.doi.org/10.1007/BF00413603`.

Tarski, A (1936a). 'O Pojęciu Wynikania Logicznego.' *Przegląd Filozoficzny*, **39**: 58–68. Page references to English translation 'On the Concept of Logical Consequence' in Tarski (1983: 409-420).

Tarski, A (1936b). 'O Ugruntowaniu Naukowej Semantyki.' *Prezeglad Filozoficzny*, **39**: 50–57. References to English translation 'The Establishment of Scientific Semantics' in Tarski (1983: 401-408).

Tarski, A (1939). 'On Undecidable Statements in Enlarged Systems of Logic and the Concept of Truth.' *Journal of Symbolic Logic*, **4(3)**: 105–112.

Tarski, A (1983). *Logic, Semantics, Metamathematics*. 2nd edn. Indianapolis, IN: Hackett.

References III

VELTMAN, F (1996). 'Defaults in Update Semantics.' *Journal of Philosophical Logic*, **25(3)**: 221–261. URL
http://www.pgrim.org/philosophersannual/xix/velt/index.htm.