Week 2: Informational teleosemantics

Mental Iconicity Seminar

Prof. Gabriel Greenberg

1.13.25

1. The problem of intentionality (and four working hypotheses)

The problem of intentionality. Content constitutes an abstraction description of the world that may be true or false; but how does content arise from the physical ingredients of natural selection, where there is no ür-interpreter?

Personal and subpersonal intentionality. We are interested in how and why mental content comes about. This potentially includes both *personal level* content that we are conscious of, and *subpersonal level* content that we are not— e.g. the contents of intermediate perceptual states or motor commands.

Working hypothesis 1: what explains one also explains the other.

I'll generally approach mental content as an objective posit, rather than a subjective observation.

Naturalism. *Naturalism about intentionality* (working hypothesis 2) is just the idea that we can explain intentional properties in naturalistic terms. This doesn't have to imply corny reduction conditions. It does imply robust explanations for how non-contentful states sometimes give rise to contentful states.

Externalism. *Internalism* is the idea that mental content might be explained purely in terms of internal descriptions, images, syntax, or inference. *Externalism* (working hypothesis 3) is the idea that mental content must be explained in part in terms of the history of interaction between internal states of the organism and external states of the environment.

Content as explanatory abstraction. Content is part of specific, relatively high, level explanation of behavior. It is part of a general level-based approach to explanation in cognitive science. We should think of contents as explanatory abstraction of brain-world interactions (working hypothesis 4), and *not* as platonic objects mysteriously tethered to the brain.

2. Informational teleosemantics (the very idea)

A brief intellectual history of informational teleosemantics. Shannon 1948 introduces the modern mathematical theory of information. Dretske 1981 proposes that we ground content in Shannon-style information. Millikan 1984 proposes to ground content in functional correlations grounded in biological function. Subsequent authors combine these ideas to create *informational teleosemantics*. Neander 2017 and Shea 2018 are the most prominent and detailed modern versions. Garson 2019 defends teleosmantics in the context of contemporary philosophy of biology. Despite the promise of the program, there remain major challenges and unanswered questions. It is still a theoretical conjecture.

What follows is my own idiosyncratic take on the project of informational teleosemantics.

Physical information is encoded by any variable state of the world that is causally downstream from another variable state of the world. Physical information is necessarily accurate. It is also ubiquitous.

Tree rings, footprints, retinas. When the causal link between one state and another is systematic and law-like, the encoded information becomes especially useful. Thus tree rings carry information about years lived; footprints carry information about the location of feet; and retinal stimulation carries information about the arrangement of objects in the world.

Functions to carry information. Organisms have many biological functions, why not functions to carry information? It would be useful for an organism to have access to information about the external world.

Content as functional information. The content of a mental state is the information which the corresponding brain state functions to carry.

Truth as success of an informational function. A mental state is *accurate* (*true*) iff the information it functions to carry is in fact carried by it, and inaccurate otherwise. Hence: if the function of state S is to carry the information that it is daytime, then S is accurate iff it is daytime and as a consequence S is tokened.

3. Some simple cases

Case study: idealized Fly detector (Fodor). Suppose an idealized frog has a neural structure N whose function is to ensure that if N=1 (at v, t, w), *there is a fly* (at v, t,w), and if N=0, *there is no fly* (at v, t,w). Further, the frog's snapping mechanism is conditioned to the value of N.

Content. Intuitively, it seems like we should get a semantics like this. For system N:

[[1]] = { *w* | there is a fly at *w* }

[[0]] = {w | there is no fly at *w*}

Metasemantics? How do we make the bridge from functional correlation to content?

Case study: System A

Organism O must refill a fluid reserve U when it is empty in order to survive and reproduce.

Semantic field A enters into two states: active and inactive.

Computational module M takes inputs from A:active, inactive and outputs motor actions M:fill, wait.

System A's informational function

The function of system A is satisfied in context c iff

for any state ■ produced by A in c:

if activation(■) = 0, then Empty(Tc);

if activation(■) = 1, then Not-empty(Tc).

4. Functions

Functions, broadly. Following Millikan we are working with a notion of teleological function that includes but is broader than biological function; it also includes learned functions. (Also called "task functions" or "proper functions.")

Stabilization. A key feature of functions is that they are **stabilized**— that is, over time, an agents disposition to fulfill that function becomes more fixed and reliable.

Wright-style analysis of function:

The function of an X is to Z iff (i) X's cause Z; (ii) the fact that (i) obtained in the past helps explain why (i) continues to obtain.

For the biological case: (ii) becomes: the fact that (i) obtained in the past promoted the survival of the organism.

For the learned case: (ii) becomes: the fact that (i) obtained in the past helps explain why the mechanism in question was reinforced.

See Wright 1973, 1976, Godfrey-Smith 1994.

Defining functions by success conditions:

Success conditions. *S* producing *R* at *w* is *successful* iff E holds at *w*.

Notation:

S:R*w* — S produces R at *w*.

[S:R*w*]=1 — the function of S producing R at *w* is successful.

[S:R*w*]=1 iff E(w). — success conditions.

5. Informational functions

Information as correlation. The idea at its simplest is that one state-type A carries information about state-type B iff whenever A occurs, B occurs, and whenever B occurs, A occurs.

Probabilistic correlation.

One-directional Correlation:

Correlational information: State type-A *carries information* about state type-B iff $\forall w \in$ N@: (A holds at *w*) $\rightarrow$ (B holds at *w*) where N is a set of normal or nearby worlds relative to @; and A is not necessary in N.

Key ideas:

Information itself can be accidental, but functional information can't.

Correlational information is perfect, but functions to carry information may fail.

Schema for informational functions:

System S functions to carry F-information about *t*.

F is a class of properties.

t is the target situation.

A target specifies a world and time, and sometimes a viewpoint and object.

I'll often simplify *t* to a world *w*.

S functions to carry information about target t ≈

S functions to produce internal states that **covary** with external states of t.

Covariation involves the counterfactually supported co-occurence of state types.

Functional covariation in the states of S may be "read off" by downstream computations.

I'll o

Functions to carry information.

In success-conditional terms: For each representation R, there is a schema of the form: [S:R*w*]= 1 iff R carries F-information about *w*.

Using the minimalist definition of correlational information: [S:R*w*]= 1 iff ∀*w'*: S:R*w'* → F(*w'*).

Gloss: *S* producing *R* at *w* is successful iff for all nearby worlds *w'*: if *S* produces *R* at *w* then *w'* is an *F*-world.

Note: imperfect correlations. a system can function to establish perfect correlation even if never achieves perfect (modal) correlation. Only the positive instances of the perfect correlation will end up reinforcing the correlational behavior.

6. Content

Semantic notation: [[*P*]]S = the content of *P* in *S*.

Option 1: accuracy conditions as success conditions. For any system *S* and representation *R* of *S*: [[*R*]]*S* = {*w* | [S:R*w*]=1 }

Objection: this makes the representation and the producer part of the accuracy conditions.

Option 2: accuracy conditions as successful correlation conditions

For any system *S* and representation *R* of *S*: [[*R*]]*S* = { *w* | *E*(*w*) } where *E* is the world-state such that if [S:R@]= 1 iff ∀*w'*: S:R*w'* → E(*w'*)

Application to the fly detector.

[[1]] = { *w* | there is a fly at *w* }

[[0]] = {w | there is no fly at *w*}

Question. What explanatory power or abstraction does defining content in this way achieve? It allows us to talk directly about the information a representation functions to carry, while abstracting away from the conditions of functional success.

7. Mental versions of I, S, and S*

System I/S/S*

Organism O must periodically refill a fluid reserve U in order to survive and reproduce.

U may contain five different volumes of liquid: 0 gallons, 1 gallon, … 4 gallons.

Semantic field I enters into one of five states of activity, measured as levels of activity 0 u, 10 u, … 40 u.

Computational module M takes inputs from I:0-40 and outputs motor actions M: add 0 gal., add 1 gal., … add 4 gal.

System I's informational function

The function of system I is satisfied in context c iff

for any state ■ produced by I in c:

volume(*Tc*) = activation(■) ■ ■ gallons.

System S's informational function

The function of system S is satisfied in context c iff

for any state ■: if S produces ■ in c, then:

if activation(■)= 20, then volume(targetc) = 0 gallons;

if activation(■)= 40, then volume(targetc) = 1 gallon;

if activation(■)= 10, then volume(targetc) = 2 gallons;

if activation(■)= 30, then volume(targetc) = 3 gallons;

if activation(■)= 0, then volume(targetc) = 4 gallons.

Is teleosemantics extensional?

It isn't clear that teleosemantics can distinguish System I from System S*, assuming we should

Don't they have the same covariational functions described in terms of external success conditions?

And past instances that supported System I would also have supported System S*.

If we want to distinguish them at the level of function, we need something more fine-grained to work with, beyond mere covariation.

8. Millikan on Intentional Icons and Semantic Rules (quotes)

Intentional icons. Intentional icons are produced by systems designed to make abstract pictures or icons that will map coincident with predetermined mapping rules to which their consumers are adjusted. When the systems in which they are embedded are functioning normally, they will picture in accordance with these rules, and they are then said to be 'true' or to be 'satisfied'.

Intentional icons and arbitrary representation. A bizarrely coded secret message from a CIA agent can be as much an 'icon' or 'picture' that maps on to a certain world affair in accordance with a definite semantic-mapping function as any bee dance, sentence, or diagram.

Semantic rules. The producer produces a sign that will be true or satisfied only if it maps on to some affair or affairs (the plural is for pushmi-pullyu icons) in the world in accordance with certain 'semantic' rules.

Semantic rules and functional success. The semantic rule associated with a descriptive intentional icon determines a condition or state of affairs that must obtain if the consumer is to perform its tasks, whatever they may be, in the normal way.

Intentional icons and semantic rules. The bee dance is, in part, a descriptive intentional icon because if the watching bees are to achieve their function of finding nectar by reacting to the dance in the normal way it needs to correspond by a certain rule to a fact about nectar location.

Note also that in the case of descriptive icons the producer's job is primarily to make an icon that corresponds by the right rule to a state of affairs.

Complex representations and rules. Mental terms are not endowed with meaning first and then used to build mental sentences. 'What makes the mental term "horse" stand for horses?' is not the place to begin. Parts and aspects of complete intentional icons represent parts and aspects of complete states of affairs only as abstracted from completed icons.

Symbols as pictures in time. These signals, taken along with their times of occurrence and sometimes with their places of occurrence, are intentional icons because variations in the times and places of occurrence correspond to variations in the times and places of the complete affairs represented.