The philosophical problem of intentionality

readings

Morgan & Piccinini, “Towards a Cognitive Neuroscience of Intentionality”

Shea, “Naturalizing Intentionality”

Optional:

Michael Rescorla, “The Computational Theory of Mind” (SEP);

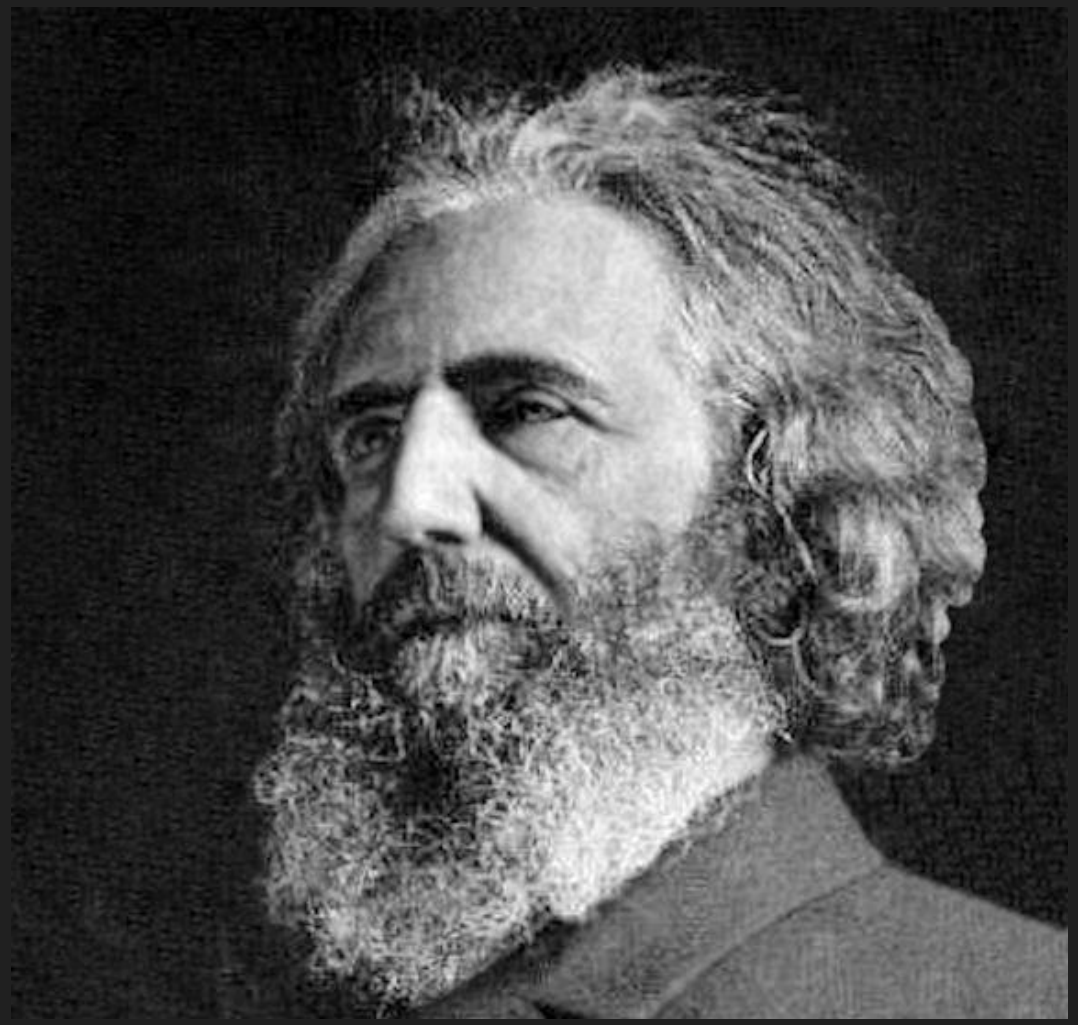
Michael Rescorla, “The Language of Thought Hypothesis” (SEP);

Neander & Schulte – “Teleological Theories of Mental Content” (SEP)

Historical notes on ‘intentionality’

Franz Brentano (re)introduced the terms ‘intentionality’ and ‘intentional directedness’ in his book, *Psychology from an Empirical Standpoint* (1872).

- the property of having a certain *subject matter*: of being *of* or *about* something.
- not to be confused with ‘intentionality’ in the more colloquial sense of purposeful. Intentions to act *do* possess the property of intentionality (they are about states of affairs one plans to bring about), but many other mental states do as well (beliefs, desires, fears, hopes, wishes, love, and so on).



Notable features of intentionality

The intentional 'relation' that holds between a mental state and what it is about (its 'object') differs in striking ways from ordinary sorts of physical relation.

- Ordinary physical relations hold regardless of how the relata are described, represented, or considered.
    - Not so with intentional relations. E.g., S may believe that Phosphorus is the brightest star in the morning sky but not believe the same about Hesperus, even though 'Phosphorus' and 'Hesperus' are names for the same thing (Venus).
- Ordinary physical relations – e.g., of being rained on, of being subject to a certain gravitational influence, of being taller than – can hold only between entities that exist.
    - Not so with intentional relations. E.g., Ponce de León travelled to Florida in search of the Fountain of Youth, believing he would find it there. What was de León thinking about?
    - See also perceptual hallucinations (e.g. MacBeth's dagger)

Intentionality in the 'formal' mode

Chisholm (1957) recast these notable features of intentionality in terms of distinctive logical profile of propositional attitude *reports:* sentences like "S believes that p." He noted that such sentences have the important property of being non-extensional ('intensional'). Specifically, attitude reports exhibit:

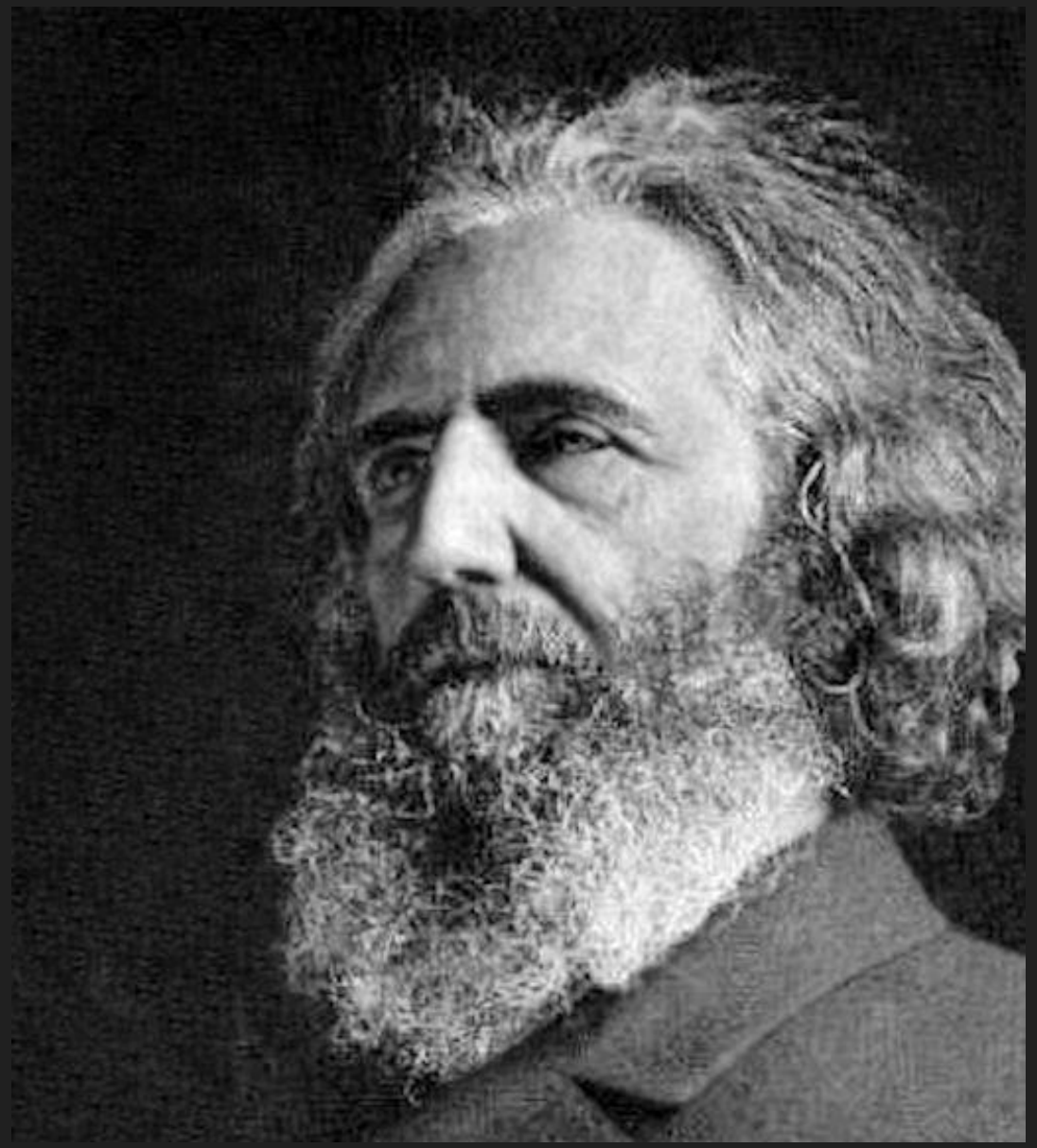
(i) failure of existential generalization (e.g., from "S believes that Santa Clause lives in the North Pole", we cannot validly infer "there exists an individual, Santa Clause, who lives in the North Pole")

(ii) non-substitutivity of co-referential terms: from "S believes that Hesperus is the first star visible after sunset", we cannot validly infer "S believes that Phosphorous is the first star visible in the night sky", even though Hesperus and Phosphorous refer to the same thing (namely, the planet Venus).

Some notable features of intentionality

We can not only think about nonexistent objects but also:

- Entities that exist outside space and time (e.g., numbers);
- future and past events;
- counterfactual scenarios (e.g., how the world might have been had events occurred otherwise than they actually did)
- entities otherwise causally isolated from us (e.g., those outside our light-cone);
- impossible states of affairs (e.g., that violate the laws of nature or perhaps even the laws of logic)

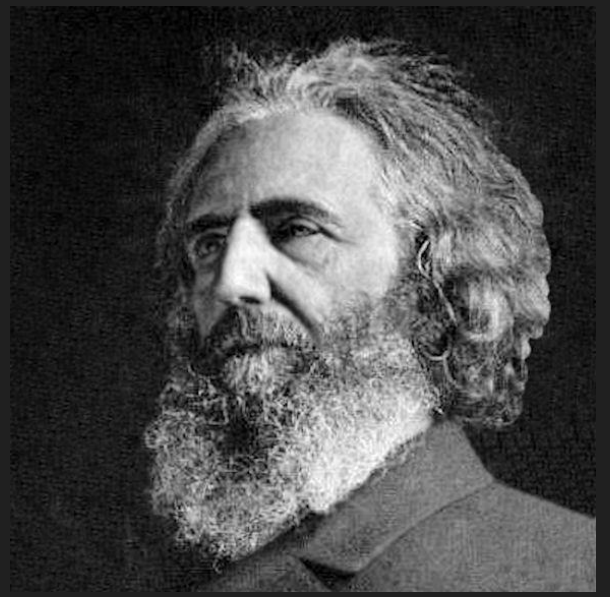


Notable features of intentionality

Even where the object of thought is real, it isn't easy to see how its being the object of thought can figure in a causal explanation of the subject's behaviour, given that semantic/intentional relations to objects appear to be non-effective. And yet, we *do* commonly explain behaviour by reference to what one was thinking.

- Does the object of one's thought exert a magnetic pull on one's brain and body when one acts with it in mind?
- Can you build a device that will alert you whenever someone thinks about you?

Questions about Brentano's thesis



Brentano claimed (and Chisholm agreed) that intentionality is the mark of the mental: that <u>all and only</u> mental states have intentionality. This is called 'Brentano's thesis.'

Is Brentano's Thesis true? In particular:

(i) Are *all* mental states about something?

(ii) Are *only* mental states about something?

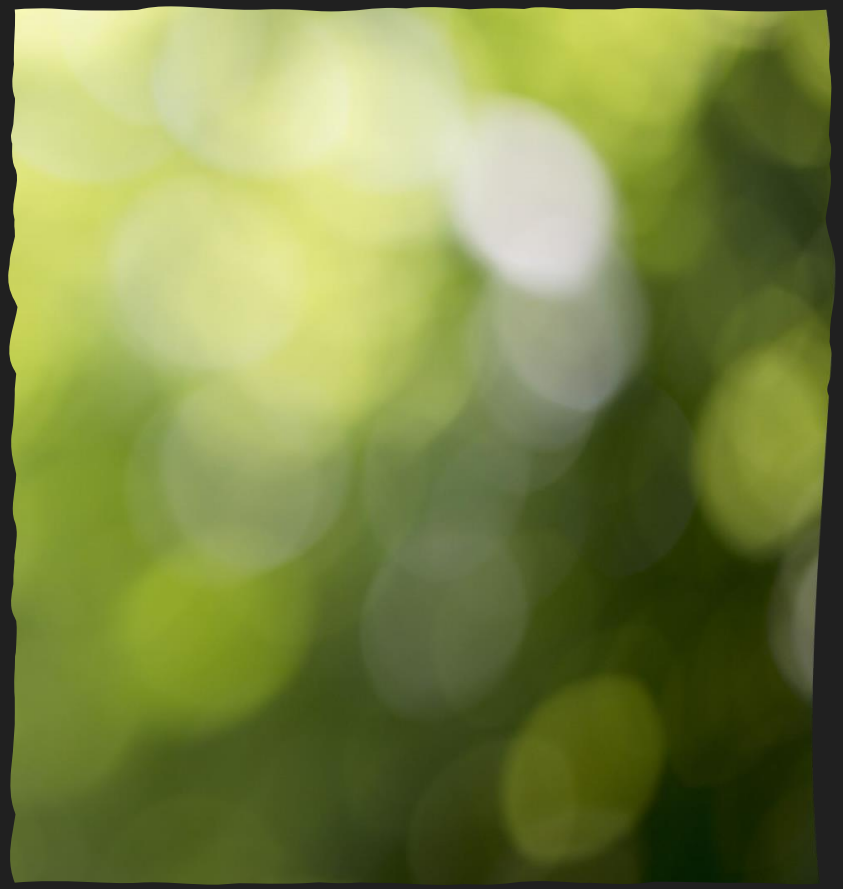
- What about public symbols, like words of natural language or pictures that depict a person?

‘Original’ vs. ‘Derived’ Intentionality

“Cognitions, perceptions, and actions *are not the only* phenomena that exhibit intentionality. A line in a song, or on a plaque, can say that Babe Ruth built Yankee Stadium; and, if so, it has essentially the same content as my belief [that Babe Ruth built Yankee Stadium]. Clearly, sentences, formulae, and public symbols of all sorts can have intentional contents, including a great deal of overlap with the possible contents of thoughts. Intentionality, however, is not all created equal. At least some outward symbols (for instance, a secret signal that you and I explicitly agree on) have their intentionality only derivatively – that is, by inheriting it from something else that has that same content already (e.g., the stipulation in our agreement). And, indeed, the latter might also have its content only derivatively, from something else again; but, obviously, that can't go on forever. Derivative intentionality, like an image in a photocopy, must derive eventually from something that is not similarly derivative; that is, at least some intentionality must be original (nonderivative). And clearly then, this original intentionality is the real metaphysical problem; for the possibility of delegating content, once there is some to delegate, is surely less puzzling than how there can be any in the first place.” (Haugeland, “Intentionality All Stars”, p. 385).

- Cf. the distinction between ‘intrinsic’ and ‘attributed’ intentionality (John Searle)

Brentano's thesis



Given the distinction between original and derived intentionality (or between intrinsic and attributed intentionality), we might tweak Brentano's thesis as the thesis that all and only mental states possess *original* or *nonderived* intentionality.

‘Direction of fit’

- All intentional mental states possess satisfaction or success conditions. And different types of intentional state can be distinguished by their characteristic standard of success (or ‘direction of fit’).
  - ‘Mind-to-world’ (e.g., beliefs and perceptions): truth, accuracy, correctness
  - ‘World-to-mind’ (e.g., desires and intentions): fulfillment, enactment, action-guidance
  - hybrid structures (e.g., ‘push-me-pull-yu’): both directions at once

Representational theory of intentionality

- Intentional mental state = a mental representation
- Intentional directedness = the representational content of a mental representation

Representational theory of intentionality

- Brief history:
  - early modern period (e.g., Locke, Hume);
    - 'Images' and 'ideas'
  - introspectionist psychology
    - Overthrown by behaviourism
  - post-behaviourist computational cognitive science
    - Intentional behaviour is the effect of internal computation, understood as the algorithmic manipulation of information-carrying representations (classically: symbols).
    - Representations are effective in virtue of their *syntactic* properties (roughly, their 'shape')
    - Physically implemented (or realized) via neural states and processes
    - No necessary connection with consciousness and introspection

Philosophical background for the representational theory

- Wilfred Sellars' myth of Jones (from *Empiricism and the Philosophy of Mind*)
  - intentional psychology as a 'folk theory' regarding the hidden causes of behaviour
  - Mental states are not directly known first-personally but postulated third-personally to predict and explain the behaviour of other agents (in the first place) and of oneself (secondarily).
- Willard Van Orman Quine's argument for indeterminacy of meaning/translation
  - underdetermination of a theory by its data
  - No purely extensional account of intentionality (cf. Chisholm)
  - Eliminativism about intentional phenomena

Interpretivism and Intentional systems theory

- Anchors our account of intentionality to the 3rd-personal perspective of an *interpreter* (Dennett; Davidson).
- Content is determined by the most charitable rationalizing interpretation of an agent's behaviour (i.e, one that maximizes the agent's intelligibility/coherence).
- Intentional states aren't unobservable 'internal' causes of behaviour but observable patterns that can be exploited in prediction and explanation.
  - In trying to arrive at a stable interpretation of an agent, one is not speculating about hidden facts but the most illuminating and charitable interpretation.
- Strongly 'externalist' about content, i.e., the content of a mental state is "wide" or constitutively dependent on features of the individual's environment.

Interpretivism and Intentional systems theory

- On some versions, the original/derived intentionality distinction is rejected (e.g., Dennett "Intentional Systems Theory", §3).
  - I.e., no distinction between *genuinely* believing that p and *merely seeming* to believe that p.
- Questions for these accounts:
  - Is this not still *anti-realism* or *instrumentalism*?
    - For Dennett's reply to this objection, see his "Real Patterns (1991). For discussion, see John Haugeland "Pattern and Being" and William Seager "Real Patterns and Surface Metaphysics"
  - What is the intentional status of *the interpreter*?
  - What (if anything) does interpretivism have to say about the mental representations posited by cognitive science?

- While interpretivism still has its proponents, most philosophers have taken a more recognizably 'realist' stance toward intentional states, regarding them as discretely identifiable internal states of the person (ala Sellars).
- Once we conceive of intentional states as unobserved internal states of the subject that serve to cause intelligent behaviour, we can ask how these putative states relate to the postulated internal states described within cognitive science.
- "Sellars and many others held that the explanatory and predictive success of folk psychology provides strong reason to think that the theory is *true*, and that intentional states really *exist*. But how, precisely, do the intentional states posited by folk psychology relate to the computationally manipulable representations posited by cognitive science?" (Moran & Piccinini, p. 125)
- One popular answer: they are (very roughly) the very same states.

"If we use "thought" in a generic sense, to cover all intentional or representational mental states (e.g., beliefs, desires, memories, perceptions, imaginings, loves, hates, and unconscious representations of features of the world), the key question is this: What obtains between a thought about something and that which it is about *in virtue of which* the former *is* a thought about the latter? What *constitutes* a mental representation representing what it represents?

One possibility is that intentionality is fundamental and inexplicable. However, this proposal does not *explain* intentionality – it abandons the possibility of explaining intentionality. Those who seek what is often called a naturalistic theory of mental content look for a way to explain how meaning could arise from nonmeaning in the natural world." (Karen Neander, p. 383).

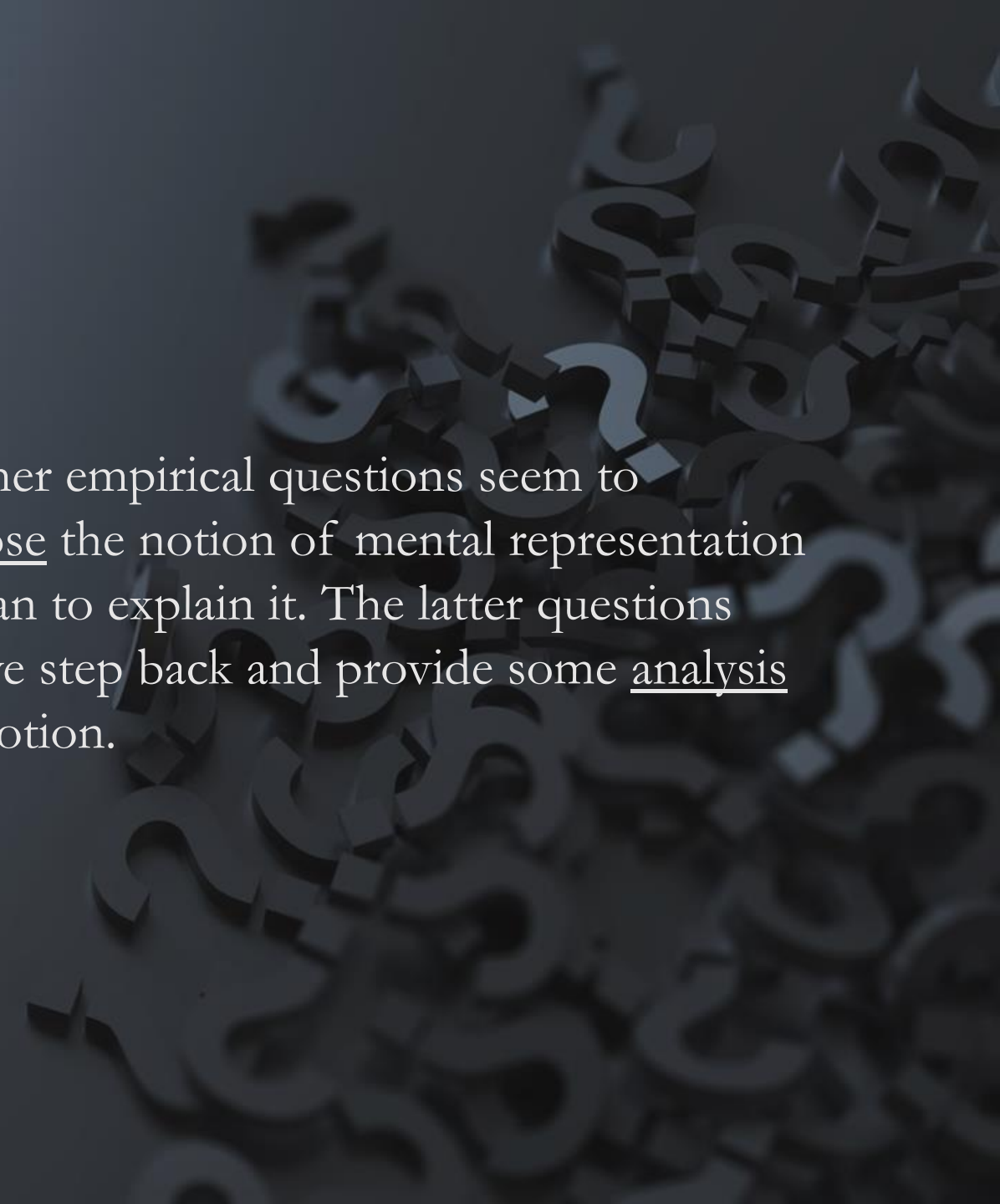
Questions about intentionality

On the one hand, there are scientific-empirical questions for cognitive science about how the brain constructs, stores, and uses mental representations to execute various computational tasks (e.g., about representational *formats*, *stages* of information-processing, patterns of informational exchange between subsystems, etc.)

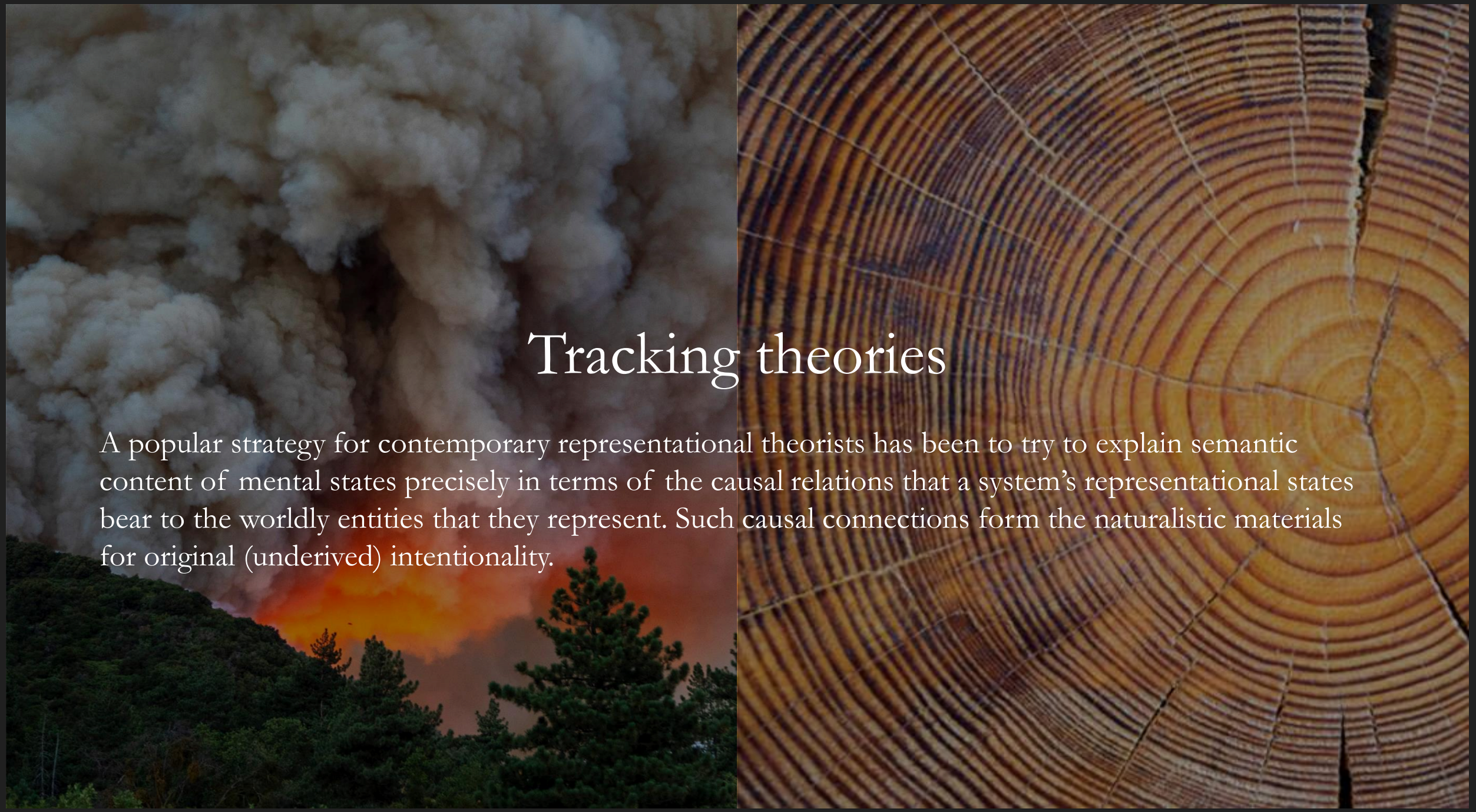
On the other, there are the conceptual questions of *what it is* to represent and how the *content* of a representation is determined (e.g., what makes Susan's current thought a thought *about her dog* rather than about, say, the neighbour's cat, New Zealand, or Pegasus?)

Questions about intentionality

The former empirical questions seem to presuppose the notion of mental representation rather than to explain it. The latter questions require we step back and provide some analysis of that notion.



Tracking theories of intentional content



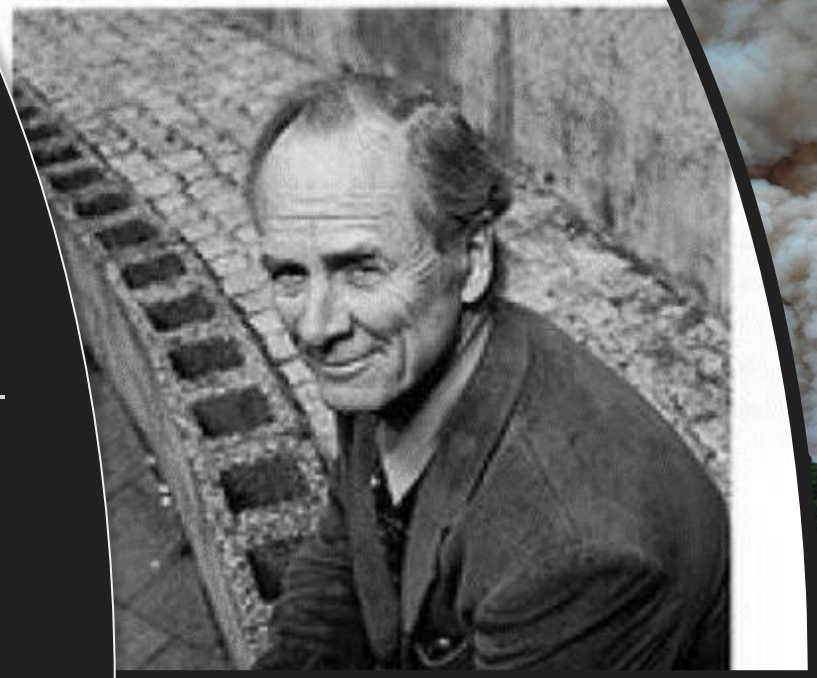
Tracking theories

A popular strategy for contemporary representational theorists has been to try to explain semantic content of mental states precisely in terms of the causal relations that a system's representational states bear to the worldly entities that they represent. Such causal connections form the naturalistic materials for original (underived) intentionality.

Tracking theories

Let's look at a representative example of such an account: Fred Dretske's 'indicator teleosemantics'.

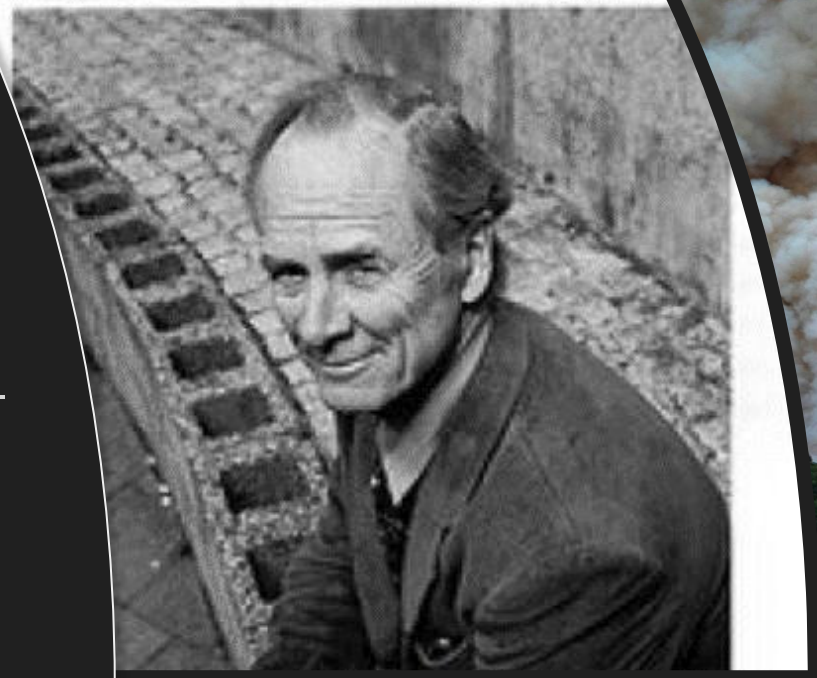
X represents Y if and only if X has the proper function of reliably indicating Y.



Tracking theories

Let's look at a representative example of such an account: Fred Dretske's 'indicator teleosemantics'.

X represents Y if and only if X has the proper function of reliably indicating Y.



Reliable indication

Non-mental examples:

- smoke reliably indicates fire
- 47 tree rings reliably indicates 47 years

In each case, the former item can be relied on to detect the presence of the latter in virtue of reliably correlating with it.

We might call such cases ‘natural meaning’ (equivalently: natural signs or natural representations) and contrast it with the ‘conventional meaning’ that public symbols possess (e.g., that ‘red’ means red or that red on a traffic light means ‘stop!’).

Reliable indication

With the notion of a natural sign, we can formulate the hypothesis that mental representations are types of physical state (e.g., patterns of neural activity) that 'indicate' or 'detect' specific types of environmental stimuli, such that tokens of the former systematically correlate with tokens of the latter.

As a detector of a particular class of entities, the detector *represents* such stimuli.

A pure indicator semantics (first-pass)

Consider the following analysis:

X represents Y if and only if X *reliably indicates* Y.

Applied to mental states, the idea would be that a mental state represents Y if and only if there is a reliable causal correlation between this type of mental state and Y.

A pure indicator semantics (first-pass)

Some have thought some such a view may be implicit within cognitive psychology, where we find invocation of various sorts of 'detectors': e.g., edge detectors in V1, motion detectors in V5/MT, facial detection in fusiform facial area.

An initial worry …

But how could reliable indication possibly get us all the fancy stuff – e.g., representation of non-existent phenomena (Santa Clause, Pegasus, the Fountain of Youth, etc.) or even of existent but *causally remote* entities (e.g., objects and events that distant in space and time)?

A reply …

Initial reply: Those are indeed hard cases. But let's try starting with relatively elementary cases in which one represents what is in fact present in their environment - e.g., perceptual representations and observational-recognitional concepts.

Once we have an account of how a system can perceive and categorize what is perceptually *present*, we will work on understanding what is involved in representing something perceptually *absent*.

Upgrading the worry into a challenge …

Okay, fine. We'll start with the simple cases. But is reliable indication even sufficient to explain the simple cases?

Two intimately related problems:

- the misrepresentation problem
- the disjunction problem

The misrepresentation problem

Notice: reliable indication *guarantees* the existence of the indicated thing.

Consider the claim that spots on one's back indicate (or mean) measles. If a patient's spots were turned out to be caused by something else (e.g., smallpox), then they never *meant* measles, but whatever it was that caused them. A doctor might mistake (misinterpret) spots as a sign of measles, but then it would be *the doctor* making the error, not the spots.

What this seems to show: indication precludes error. So, if indication is necessary for representation, then X cannot represent Y unless Y *really is* present.

The misrepresentation problem

Why is this a problem? Because even the most basic representations do not guarantee the presence of what they represent.

E.g., it is possible to mentally represent a sheep even when no sheep are present. (For example, you might mistake a goat on a dark night for a sheep).

A pure indicator semantics seems to preclude this possibility. But (many have thought) this makes it unsuited to explain representation: representation implies the possibility of *misrepresentation*.

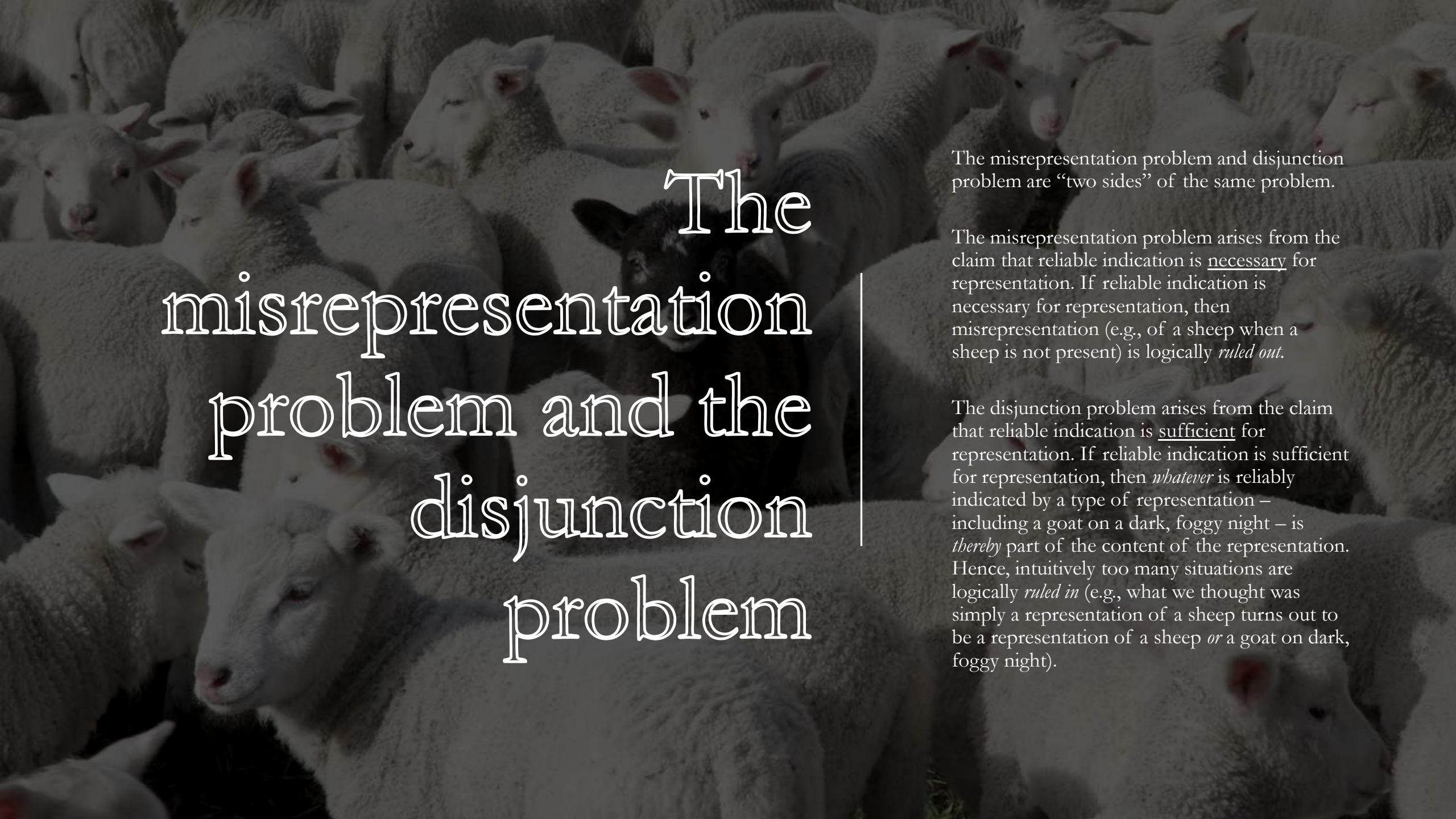

Goat
VS
Sheep

The “disjunction problem”

“Suppose that I am able to recognise sheep – I am able to perceive sheep when sheep are around. My perceptions of sheep are representations of some sort – call them ‘S-representations’ for short – and they are reliable indicators of sheep, and the theory therefore says that they represent sheep. So far so good.

But suppose too that, in certain circumstances – say, at a distance, in bad light – I am unable to distinguish sheep from goats. And suppose that this connection is quite systematic: there is a reliable connection between goats-in-certain-circumstances and sheep perceptions. I have an S-representation when I see a goat. This looks like a clear case of misrepresentation: my S-representation misrepresents a goat as a sheep. But, if my S-representations are reliable indicators of goats-in-certain-circumstances, then why shouldn’t we say instead that they represent goats-in-certain-circumstances as well as sheep? Indeed, surely the indication theory will *have* to say something like this, as reliable indication alone is supposed to be the source of representation.

The problem, then, is that both sheep and goats-in-certain-circumstances are reliably indicated by S-representations. So it looks like we should say that an S-representation represents that either a sheep is present *or* a goat-in-certain-circumstances is present. The content of the representation, then, should be *sheep or goat-in-certain-circumstances*. This is called the ‘disjunction problem’ because logicians call the linking of two or more terms with an ‘or’ a *disjunction*.” (Crane 2003, p. 179)



The misrepresentation problem and the disjunction problem

The misrepresentation problem and disjunction problem are "two sides" of the same problem.

The misrepresentation problem arises from the claim that reliable indication is necessary for representation. If reliable indication is necessary for representation, then misrepresentation (e.g., of a sheep when a sheep is not present) is logically *ruled out*.

The disjunction problem arises from the claim that reliable indication is sufficient for representation. If reliable indication is sufficient for representation, then *whatever* is reliably indicated by a type of representation – including a goat on a dark, foggy night – is *thereby* part of the content of the representation. Hence, intuitively too many situations are logically *ruled in* (e.g., what we thought was simply a representation of a sheep turns out to be a representation of a sheep *or* a goat on dark, foggy night).

The misrepresentation problem and the disjunction problem

The moral: pure indicator semantics seems to deliver the intuitively wrong answer to the question of what a given type of representation represents in certain circumstances.

How to respond?

Philosophers have explored various options for overcoming this challenge to indicator semantics.

Here, we'll continue with Dretske's proposal, which (recall) included a second component: namely the notion of 'proper function'.

Proper functions to the rescue?

Proper functions (*a.k.a.*, "teleofunctions") are standardly attributed to biological mechanisms, and teleological (or teleofunctional) explanations are explanations in terms of such teleofunctions.

e.g., proper function of a heart is to pump blood (rather than to cause a gentle thumping noise). That is what hearts are *for*.

Note: unlike the notion of "function" that functionalists operate with, the notion of proper function is explicitly *teleological*: it does not say how an entity actually behaves, but how it is *supposed* to behave. Something can have a proper function, even if it not exercising that function, or has even lost the capacity to do so. In such cases, the entity is *malfunctioning*.

Proper functions to the rescue?

Despite their prevalence in biology, the nature of proper functions is a contested topic in philosophy of biology.

According to an influential family of "etiological" accounts, the function of an item, *x*, is to *Z* in *C*, if and only if *x*s were selected for *Z*-ing in *C* by a natural process of selection (Neander p. 386).

Proper functions to the rescue?

A selective process could be one of explicit design (as in the case of artefacts), or natural (e.g., the product of evolution via natural selection). In either case, we appeal to a thing's selective history to explain why those items exist, and to distinguish the item's proper function from its other properties and effects. Roughly, an item's proper function is the effect for which the phenotypic trait was selected.

Proper functions to the rescue?

Proper functions to the rescue?

The S-representation is a representation of a sheep (rather than of a sheep *or* a goat on a dark, foggy nights) because S-representations have *the proper function* of indicating the presence of sheep (not sheep or goats on dark, foggy nights).

Further, it is possible for S-representations to sometimes malfunction – e.g., to 'misfire' in the presence of a goat on a dark, foggy night. Even when accidentally triggered by a non-sheep, the S-representation represents whatever triggers it as a sheep, since that is what it has the proper function of indicating.

We seem to have avoided the misrepresentation problem and disjunction problem!

And, in doing so, we have also shown how it is possible for a physical representation to represent something in its absence. (A modest but significant first step toward naturalizing intentionality).

Or have we?

Challenges

- the ‘functional indeterminacy problem’
- etiological theories of proper function generate some weird consequences in hypothetical scenarios (e.g., Swampman) and to suggest that content may be explanatorily irrelevant to how an agent behaves.
- Some doubt whether such a minimal notion of ‘tracking’ is sufficient for mental representation.
    - Representation seems not merely a matter of tracking but of tracking when separated from something (of using something as a ‘stand-in’ for the real thing) (cf. Haugeland).
    - Representation seems not merely a matter of tracking, but of tracking *objects* (as opposed to something merely *subjective*), suggesting a self-other distinction (cf. Burge).

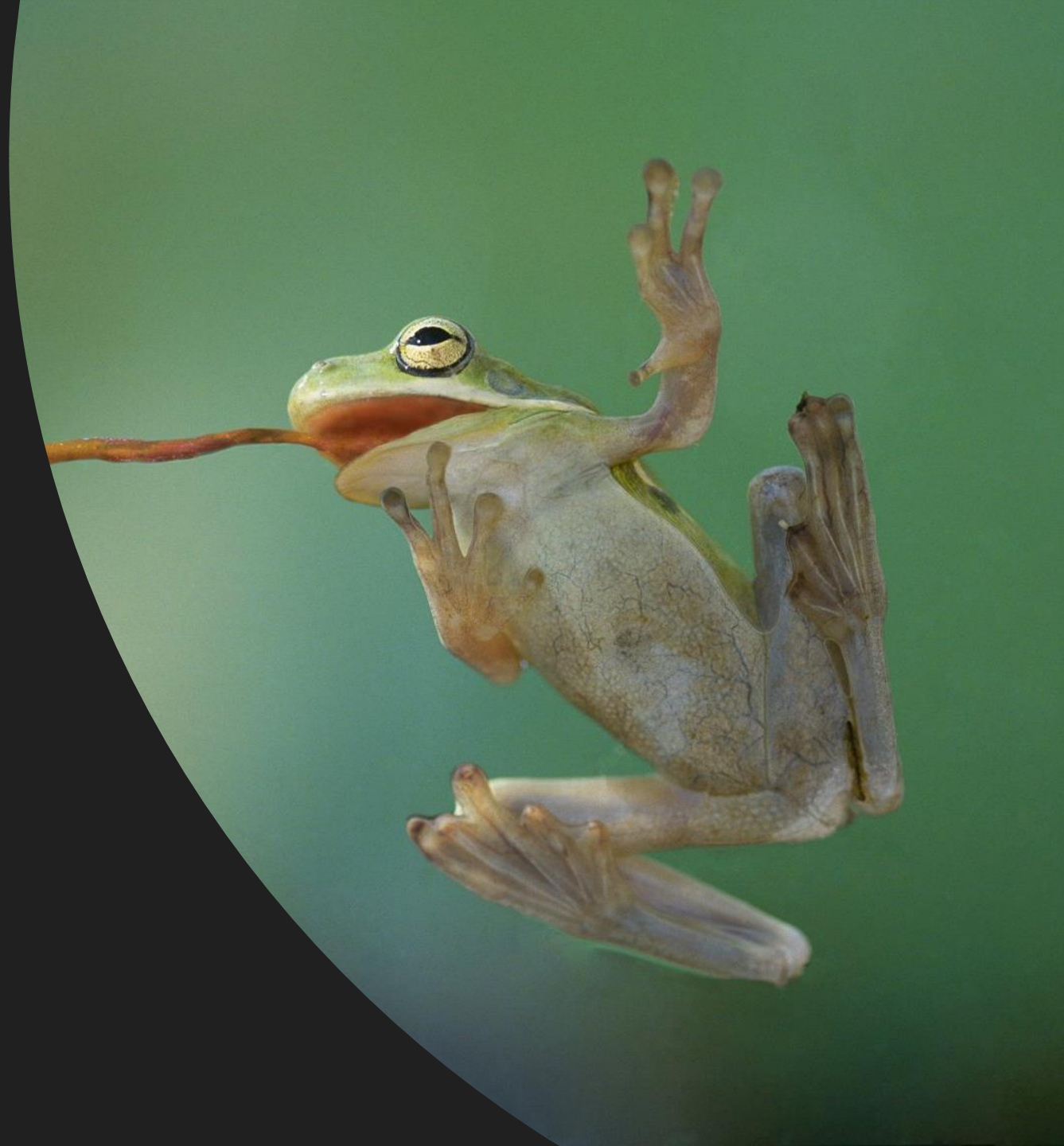
The functional indeterminacy problem

The functional indeterminacy problem

This frog is visually sensitive to the presence of flies. When it sees one, it snaps its tongue out and grabs it.

On an indicator teleosemantic account, the frog possesses a visual representation whose proper function is to reliably detect …

- A fly?
- Something small, dark, and moving?
- A certain pattern of retinal stimulation?
- food?
- Something untranslatable into human language?



Another example

"Some marine bacteria have internal magnets (called magnetosomes) that function like compass needles, aligning themselves (and as a result, the bacteria) parallel to the earth's magnetic field. Since these magnetic lines incline downwards (towards geomagnetic north) in the northern hemisphere (upwards in the southern hemisphere), bacteria in the northern hemisphere . . . propel themselves towards geomagnetic north. The survival value of magnetotaxis (as this sensory mechanism is called) is not obvious, but it is reasonable to suppose that it functions so as to enable the bacteria to avoid surface water. Since these organisms are capable of living only in the absence of oxygen, movement towards geomagnetic north will take the bacteria away from oxygen-rich surface water and towards the comparatively oxygen-free sediment at the bottom" (Dretske)

What is the proper function of the bacterium's magnetosome? Is it to propel the bacterium to *geomagnetic north*, to *the absence of oxygen*, to *safety*, or all of the above?

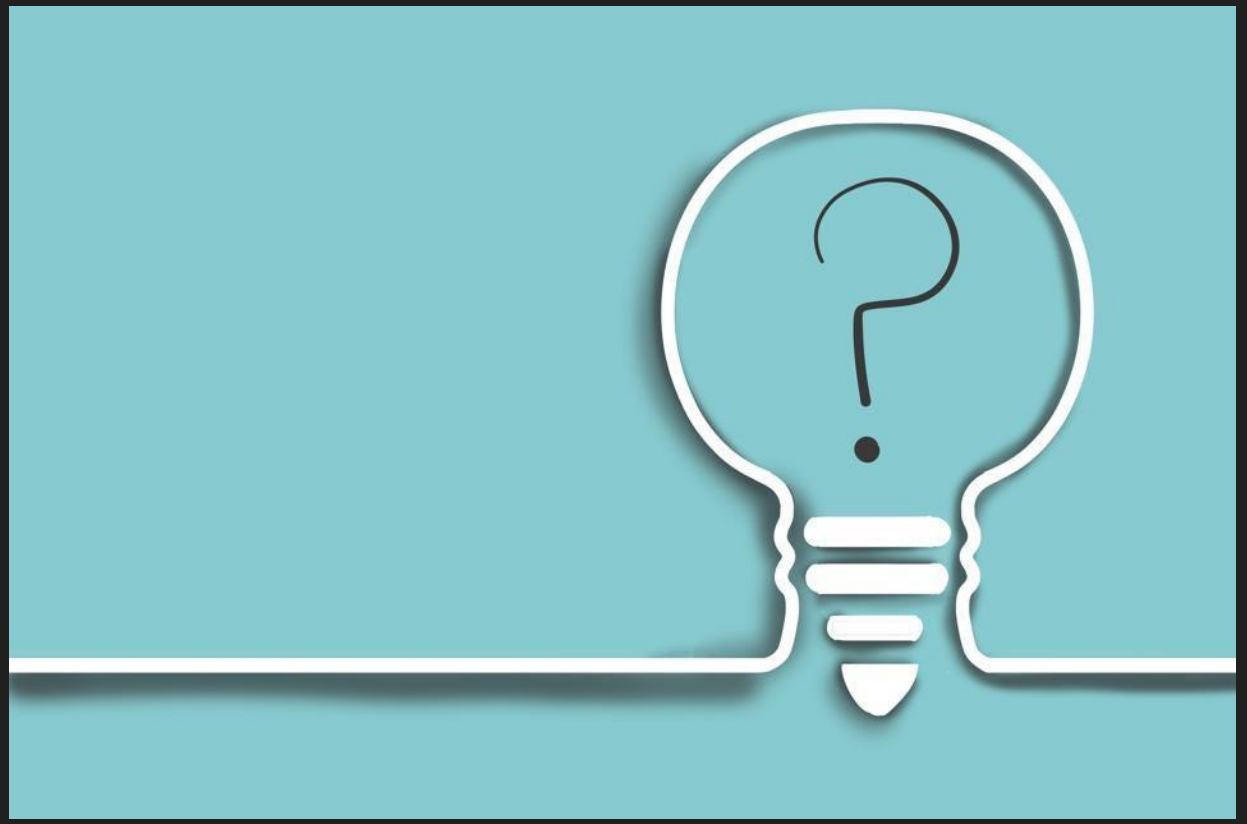
This "functional indeterminacy" is reminiscent of the disjunctive problem discussed last time, in that it introduces a wide disjunction of admissible contents.

Another problem …

Swampman:

A being molecule-for-molecule identical to you, but who came into existence by complete accident (and so has no evolutionary history)

But without an evolutionary history, there can be no teleofunctions and therefore no representational content …



Introducing consciousness

Concepts of consciousness: creature consciousness, state consciousness, and levels

Creature consciousness: consciousness as a property of a creature/organism/person. E.g., Mary is conscious, and so is her dog.

State consciousness: consciousness as a property of a mental state. E.g., Mary's experience of her dog is conscious rather than unconscious.

(Note: for some theorists, this is a merely conceptual distinction, with state consciousness being explanatorily more fundamental than creature consciousness. On this way of thinking, Mary is creature conscious if and only if Mary is in some conscious state or other. Other theorists use 'creature consciousness' to denote a distinct aspect or dimension of a conscious experience pertaining to its *global* character rather than its specific 'content.' For example, in saying that Mary is conscious, we may imply that she is awake and alert as opposed to asleep and dreaming, delirious, hypnotized, in a psychedelic state, etc. Such 'global' differences are sometimes marked with the terminology of 'levels' or 'modes' of consciousness)

Concepts of consciousness: Varieties of (state) consciousness

self-consciousness: a mental state you are conscious of yourself being in; consciousness of a mental state *as such*

access consciousness: a metal state available for use in the global control of action (explicit reasoning, verbal report, intentional action, etc.).

affective consciousness: a mental state that possesses a hedonic tone or affective 'valence' (positive, negative, neutral); cf. sentience

phenomenal consciousness: a mental state that there is something it is like for the subject to be in. The phenomenal character of such a state is how it 'feels' or 'what it is like.'

- main locus of philosophical controversy

Representationalist theories of phenomenal consciousness

- ('Reductive') representationalists about phenomenal consciousness aim to give a two step reductive explanation of phenomenal consciousness:
  - Step #1: reduce the phenomenal character of a conscious mental state to representational content;
  - Step #2: reduce the representational content of a mental state to tracking (e.g., informational and etiological) relations to the environment.

Representationalist theories of phenomenal consciousness

- An initial worry about this proposed reduction:
  - While representational content *might* be adequate to account for the 'qualitative character' of a phenomenally conscious state (though this is by no means uncontroversial), it seems clearly inadequate to account for the 'subjective character' of a phenomenally conscious state
  - Qualitative character is the phenomenal character that distinguishes one conscious experience from some other phenomenally conscious state.
  - Subjective character is the phenomenal character that all conscious experiences share in virtue of which they qualify as phenomenally conscious rather than unconscious.

Accounting for subjective character reductively

A basic fault-line between representationalist theories vis-à-vis ‘subjective character’

- First-order theories (e.g., ‘PANIC’, Global Workspace Theory)
  - Reduction of phenomenal consciousness to a functional notion such as access consciousness
- Higher-order theories (e.g., Higher Order Thought Theory)
  - Reduction of phenomenal consciousness to meta-representation

Accounting for subjective character reductively

- A major disagreement between FO and HO theories concerns the basic phenomenological structure of consciousness: transparency (world-directedness) vs. transitivity (self-directedness or reflexivity).