The Role of Representation in Computational Models

Frances Egan
Department of Philosophy
Center for Cognitive Science
Rutgers University
USA

Aims of Talk

- To explicate the role that representation plays in the sciences that aim to explain our cognitive capacities – in particular, computational cognitive psychology and computational neuroscience.
- To argue that neither a realist nor a fictionalist interpretation of representational talk is appropriate. I call the right interpretation *deflationary.*

Preliminaries: What are Representations?

- Central examples: utterances, written text, maps, photographs, drawings…

Preliminaries: What are Representations?

- Central examples: utterances, written text, maps, photographs, drawings…
- *Vehicle* (physically realized, has causal powers) vs. *Content* (meaning, information conveyed)

Preliminaries: What are Representations?

- Central examples: utterances, written text, maps, photographs, drawings…
- *Vehicle* (physically realized, has causal powers) vs. *Content* (meaning, information conveyed)
- Capable of *mis*representing

Mental Representations

Computational neuroscientists use representational talk, but should it be taken *literally*?

Do computational models really posit (internal) representations?

Mental Representations

Two obvious possibilities:

- Robust realism – representational talk is to be taken *literally*: mental representations are as real as written texts, maps, etc.
- Fictionalism – representational talk is just a useful fiction.

MENTAL REPRESENTATIONS

Two obvious possibilities:

- Robust realism – representational talk is to be taken *literally*: mental representations are as real as written texts, maps, etc.
- Fictionalism – representational talk is just a useful fiction.

A third possibility: Mental representations are (real) states/structures that are given a representational *gloss*.

Computational Models

One motivation for fictionalism – computational descriptions are just *metaphors:*

- Indeed some people go so far as to argue that the (human) brain is not just analogous to a computer in a strictly metaphorical sense, but that it actually is a computer in the sense that it takes in input, processes it in various ways, and then produces a specific output…. our use of the computer metaphor is so familiar and comfortable that we sometimes forget that we are dealing only with a metaphor and that there may be other, equally interesting (and perhaps more appropriate) ways to think about brains and nervous systems and what they do. (Louise Barrett, *Beyond the Brain: How Body and Environment Shape Animal and Human Minds*)

Computational Models

- But if the purpose is to *explain* a cognitive capacity, a metaphor won't do.
- Generally, the metaphor view is mistaken as an account of computational models. A computational model that characterizes a neural system as computing integration or vector subtraction purports to literally describe neural processes. The brain computes the specified function in the same sense as a hand calculator or iPad does.

Computational Models

- Computational characterization involves *idealization*, but a computational description of a neural system involves no more idealization than does a computational description of any other physical system.

Computational Models

- Computational characterization involves *idealization*, but a computational description of a neural system involves no more idealization than does a computational description of any other physical system.
- Despite the idealization, the computational characterization is intended to be literally true of both the brain and the calculator, and the explanatory role of these models depends on this. If it is just a metaphor there is no explanation.

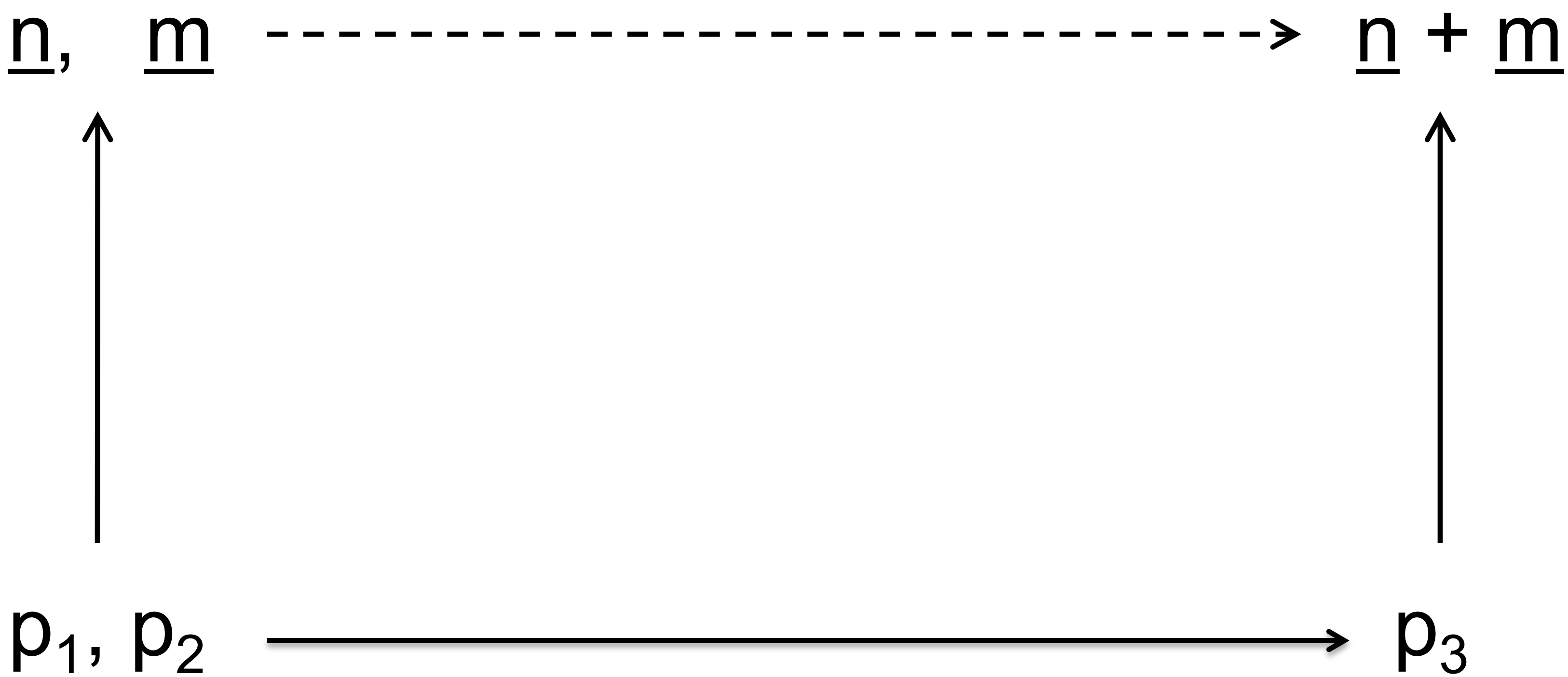
Computational Models

- At an early stage the model might be just a hypothesis, without much empirical confirmation. We might not know how it is realized in the brain; it might be the merest of ‘mechanism sketches’. But it is a hypothesis about what the brain is actually doing, so a literal interpretation is appropriate.

Computational Models

- At an early stage the model might be just a hypothesis, without much empirical confirmation. We might not know how it is realized in the brain; it might be the merest of ‘mechanism sketches’. But it is still a hypothesis about what the brain is actually doing, so a literal interpretation is appropriate.
- But what is it for a physical system to compute a function?

Example – Addition



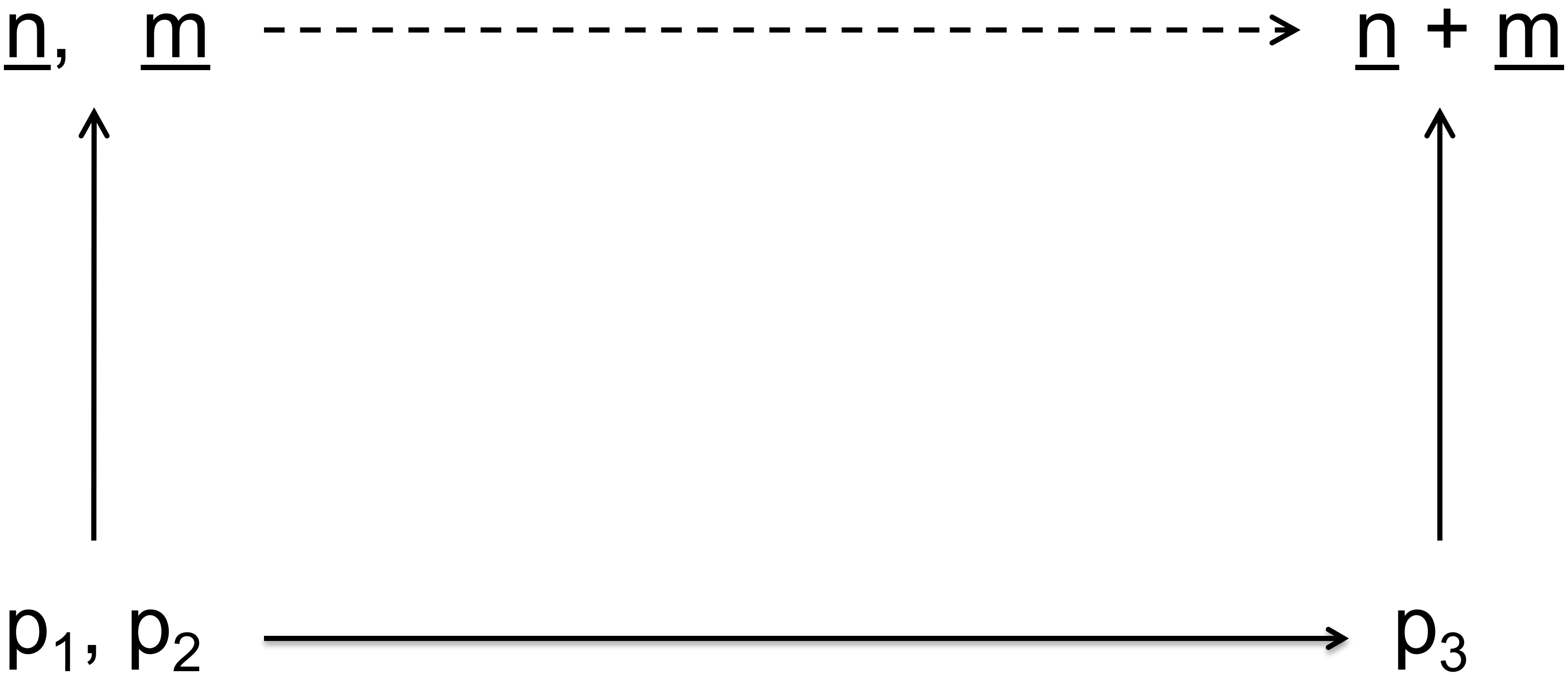
Computational Models

- Computational descriptions are *abstract*, but they are not, or needn't be, *fictions*. A computational description is *an abstract characterization of the causal organization underlying a cognitive capacity.*

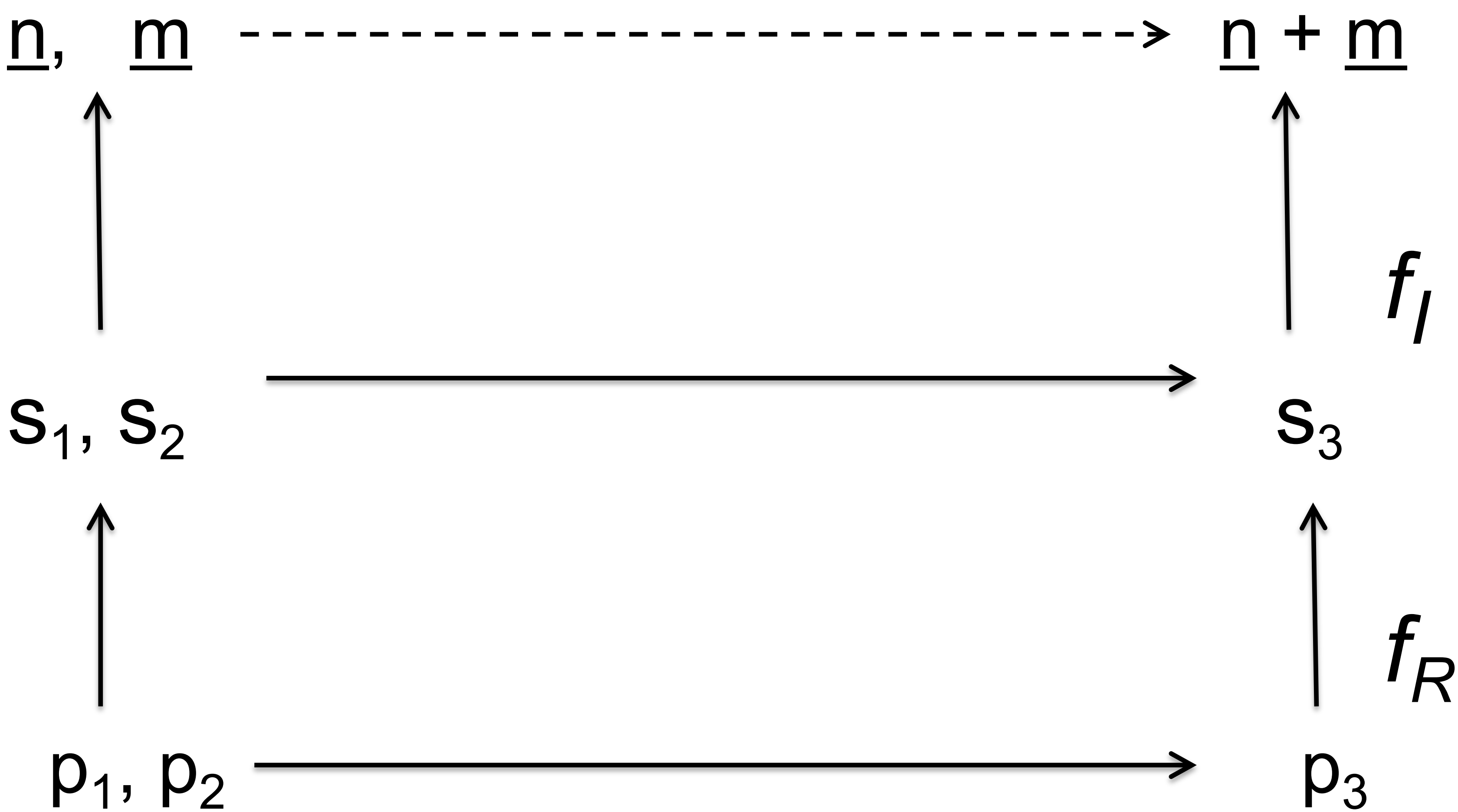
Computational Models

- Computational models characterize the causal organization of the system by specifying the function (in the mathematical sense) that the system computes, and then, in a fully articulated computational model of a human cognitive capacity, specifies the neural states that realize the arguments and values of the function.

Example – Addition



EXAMPLE – ADDITION



Computational Models

Two mappings:

- A *realization function* (f_R), which specifies the vehicles of representation
- an *interpretation function* (f_I) which specifies their content.

Computational Models

- We should be liberal about what counts as a *vehicle*. They might be discrete *symbols*, but they could also be characteristic patterns of activation of a network. The realization function (f_R) will abstract away from some properties of the realizing physical (neural) states and group them together by their role in cognitive processing.

Computational Models

- A computational theory takes states of the system to literally represent the arguments and values of the computed function. I call these *mathematical contents*, and we should be straightforwardly realist about them. So *realism* about both vehicles and contents is appropriate here.

Computational Models

More examples:

- Perceptual systems compute *smoothing functions*
- The motor control system computes *vector subtraction* functions
- The oculomotor system computes *integration* functions

These mechanisms compute well-defined mathematical functions and their input and output states represent the arguments and values of the function.

Mental Representations

- Theorists talk of the visual system representing edges in the world and of ‘place cells’ in the rat brain representing locations in the local space, i.e. they talk about internal states representing external, distal objects and properties. These contents are sometimes described as *cognitive contents*, because they are the sorts of contents that typically get ascribed in accounts of cognitive capacities.

Mental Representations

- Theorists talk of the visual system representing edges in the world and of ‘place cells’ in the rat brain representing locations in the local space, i.e. they talk about internal states representing external, distal objects and properties. These contents are sometimes described as *cognitive contents*, because they are the sorts of contents that typically get ascribed in accounts of cognitive capacities.
- How are cognitive contents determined?

Mental Representations

- *Naturalistic* proposals for cognitive contents: the relation between an internal structure and what it is about (the *representation relation*) must be specifiable without reference to *meanings* or the *intentions* of theorists.
- The representation relation is assumed to be a kind of *tracking* relation.

Mental Representations

- *Causal-based accounts* (of cognitive contents) – an internal structure means *X* if it is caused by an X.

Mental Representations

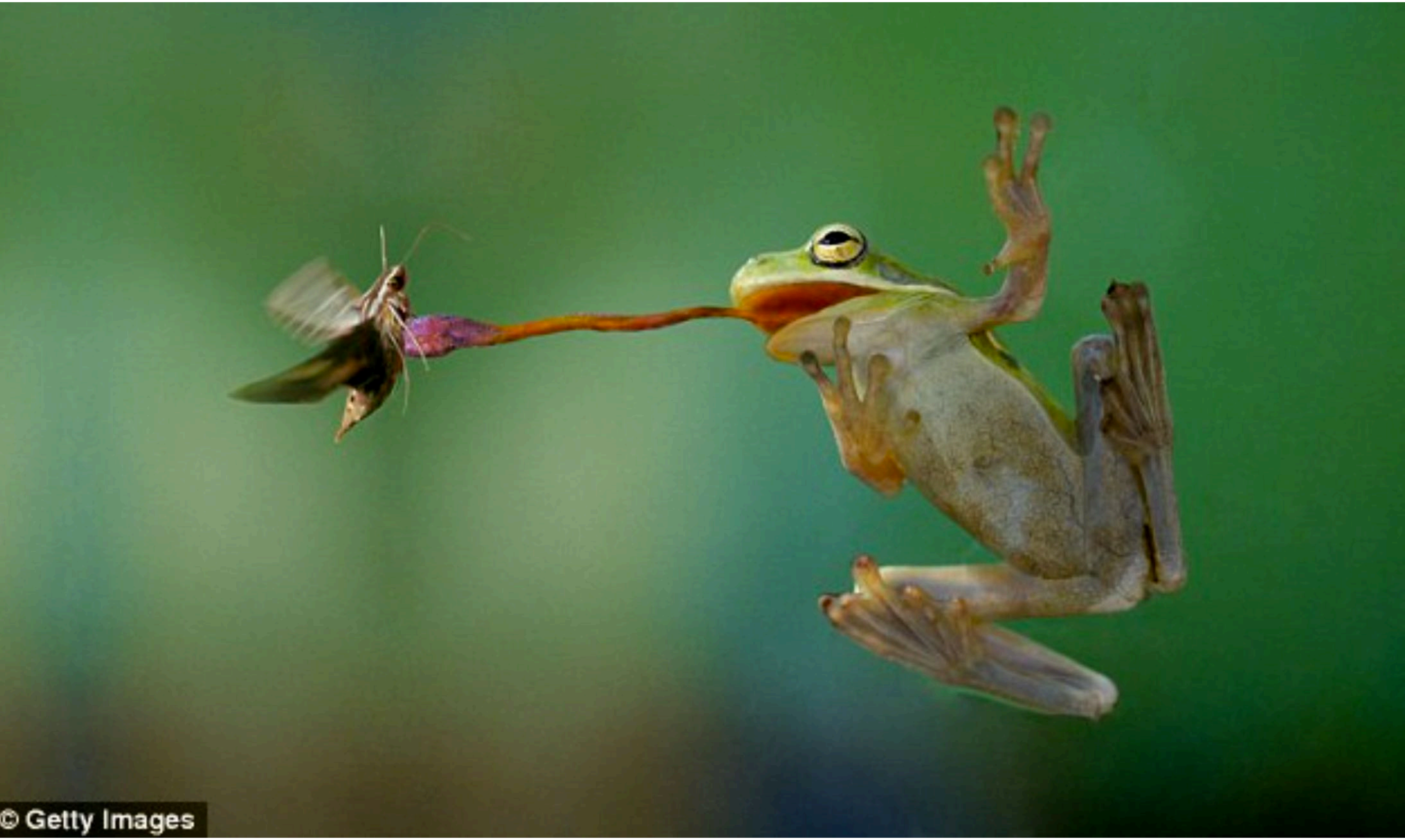
- *Causal-based accounts* (of cognitive contents) – an internal structure means *X* if it is caused by an X.
- *Teleological accounts* – an internal structure means *X* if it has the biological function of indicating Xs.

Mental Representations

- *Causal-based accounts* (of cognitive contents) – an internal structure means *X* if it is caused by an X.
- *Teleological accounts* – an internal structure means *X* if it has the biological function of indicating Xs.
- *Structural similarity (homomorphism) accounts* – the type of relation that holds between a map and the objects and properties it represents.

Mental Representations

- A problem for all naturalistic accounts of cognitive contents: *indeterminacy* – whatever the proposed tracking relation, it fails to pick out a unique external condition. The internal structure will have multiple different meanings.


© Getty Images

Mental Representations

Candidates for content:

- *fly*
- *food*
- *small dark moving object*

Mental Representations

- Response: So what… why does indeterminacy matter? Doesn't the internal structure represent *all* of these things?

Mental Representations

- Response: So what… why does indeterminacy matter? Doesn’t the internal structure represent all of these things?
- It matters because if the structure doesn’t have determinate content then it isn’t a representation. Recall that a characteristic feature of representations is that they can *misrepresent*, but if the structure doesn’t have a determinate content then it is impossible to draw a line between a correct application and an error.

Mental Representations

Candidates for content:

- *fly*
- *food*
- *small dark moving object*

Mental Representations

- But there is something right about this response: any of these contents might plausibly be attributed, just *not all at once*. So how to decide between them?

Mental Representations

- But there is something right about this response: any of these contents might plausibly be attributed, just *not all at once*. So how to decide between them?
- My proposal: we get determinate content only by appealing to the *theorist's explanatory aims and goals.* The justification for selecting one content over the others is *pragmatic*.

Mental Representations

Candidates for content:

- *fly*
- *food*
- *small dark moving object*

Mental Representations

- The general point is that indeterminacy of content is resolved only by reference to specific explanatory interests, theorists' goals and intentions (most plausibly, *collective* goals and intentions of scientific communities). And once we have determinate content we have a way of partitioning the system's responses into correct applications and mistakes (i.e. misrepresentations).

Mental Representations

- Objection: Whether the brain genuinely represents or not – whether it is doing something *enough like* our central cases of representation to count as representation – is presumably an *objective* matter, not something to be determined by reference to the goals and pragmatic choices of theorists.

Mental Representations

- Response: Despite theorists' representational talk, computational theories do not posit representations with determinate (cognitive) content. What they posit are structures that play causal roles in cognitive processes. It's a purely objective matter whether there are such structures.

Mental Representations

- Some of these structures are assigned content in an *explanatory gloss* that accompanies the theory. This assigned content – and so the pragmatic elements that determine this content – is not in the theory, but is ‘quarantined’ in the explanatory gloss.

Mental Representations

- I am appealing to a distinction between what we might call *the computational theory proper* and an *explanatory gloss* that accompanies it. Determinate (cognitive) content – a necessary condition for a structure to count as a representation – is in the gloss.

Theory vs. Explanatory Gloss

The *theory proper* specifies:

1) The mathematical function(s) computed by the mechanism,
2) the algorithms involved in the computation of the function(s),
3) the structures that the algorithm maintains,
4) the computational processes defined over these structures, and
5) general facts about the organism's normal environment that explain why computing the specified mathematical function(s) suffices, in context, for the successful exercise of the cognitive capacity to be explained.

Theory vs. Explanatory Gloss

3) the structures that the algorithm maintains
4) the computational processes defined over these structures

The structures specified by (3) are glossed as *representations* (i.e. they are assigned contents in the gloss). And the causal processes specified by (4) are glossed as *representational processes*.

Theory vs. Explanatory Gloss

Examples of explanatory glosses in science:

- Bohr *glossed* the atom as a planetary system
- Maxwell *glossed* Faraday's lines of force as the movement of fluid through tubes

A *representational gloss* construes internal structures as referring to (about) external objects and properties (e.g a structure might mean *edge)*.

Functions of a Representational Gloss

(1) A cognitive theory aims to explain a *capacity*, a *competence*. That's not a normatively neutral way of characterizing a class of behaviors – it presumes the behavior is *correct*, *successful*. A representational gloss of the posited mechanisms allows us to see an individual instance of behavior as *correct* or as a *mistake,* as accurately representing, or *mis*representing, its target (e.g. an edge in the scene).

Functions of a Representational Gloss

(2) The gloss characterizes internal structures in a way that makes clear their causal role in a process that typically extends into the environment. The content ascription *selects* what is salient in a complex causal process.

Functions of a Representational Gloss

- Example: ‘place cells’ in the rat hippocampus

Functions of a Representational Gloss

- Example: ‘place cells’ in the rat hippocampus
- What is important is specifying the conditions to which the cell’s firing is responsive and determining its role in controlling subsequent behavior.

Functions of a Representational Gloss

- Talk of the cell's firing *representing* its distal stimulus conditions is a convenience that adds nothing of theoretical significance.
- The primary function of content ascription here is characterizing internal structures and states in a way that makes clear their causal role in a cognitive process.

Functions of a Representational Gloss

(3) The gloss can also serve as a temporary *placeholder* for an incompletely developed computational theory – prior to the specification of the structures and processes that make up the theory proper – and guide the investigation into the mechanisms underlying the capacity.

Functions of a Representational Gloss

(3) The gloss can also serve as a temporary *placeholder* for an incompletely developed computational theory – prior to the specification of the structures and processes that make up the theory proper – and guide the investigation into the mechanisms underlying the capacity.

At this early stage of theory development, *fictionalism* is the appropriate attitude to take toward representational talk.

Functions of a Representational Gloss

- But fictionalism about representational talk isn't generally correct. A fully articulated theory posits structures that play causal roles in cognitive processes. These structures really exist. If they are assigned cognitive content in a representational gloss, then they are *glossed* as representations, but they are not fictions.

Functions of a Representational Gloss

- But fictionalism about representational talk isn't generally correct. A fully articulated theory posits structures that play causal roles in cognitive processes. These structures really exist. If they are assigned cognitive content in a representational gloss, then they are *glossed* as representations, but they are not fictions.
- This is a 'deflationary' interpretation of representation.

SUMMARY

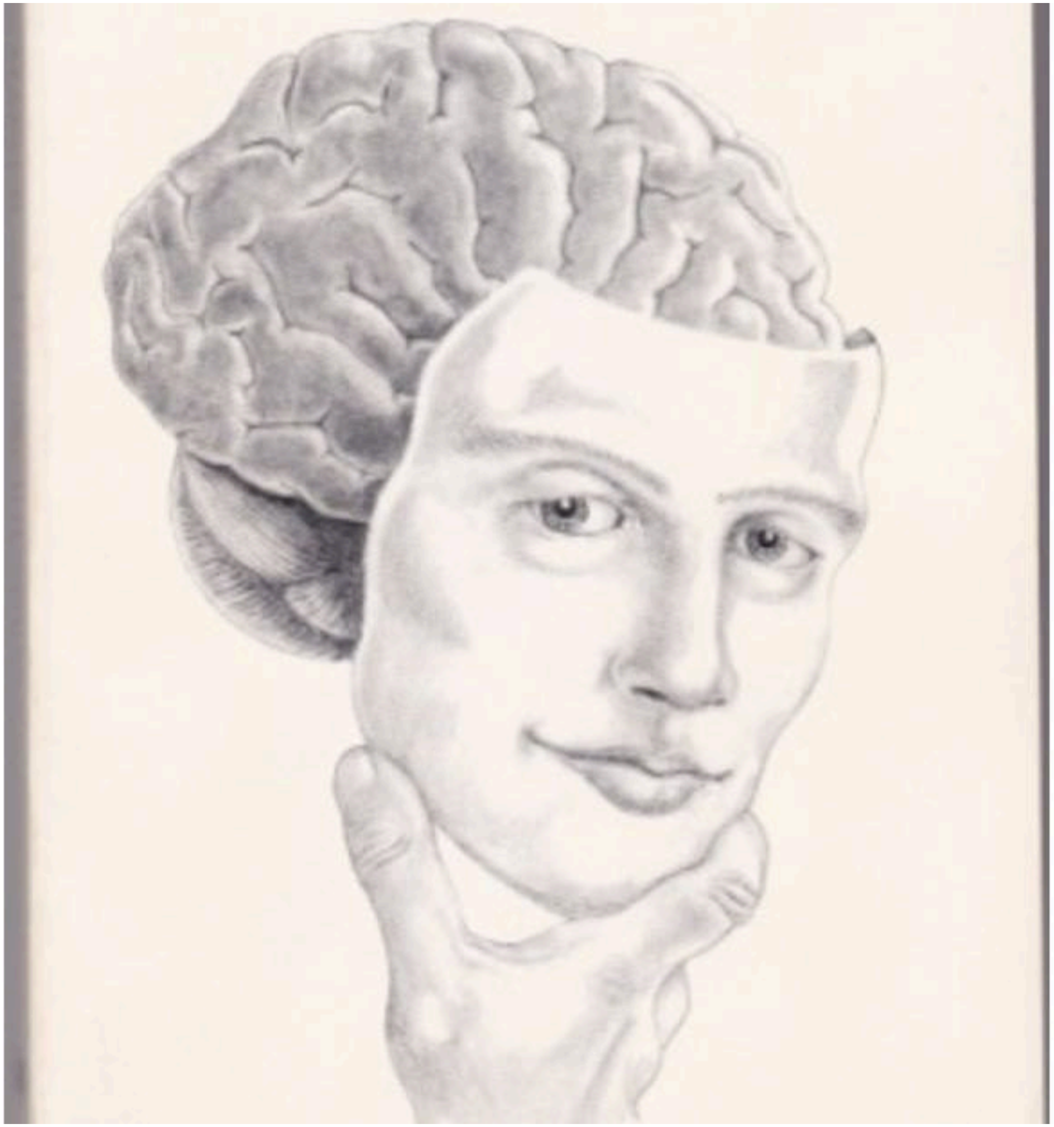
- Contents play complex heuristic roles in computational theories of cognitive capacities. They are for the benefit of theorists and students of cognition – they make intelligible the account of the mechanisms specified by the theory.

Summary

- Attributing determinate contents (in the representational gloss) allows us to see the organism's activity as accurate representation of the world and (occasionally) as *mis*representation.

Summary

- Normative notions such as *correctness* and *mistake* are confined to the representational gloss. It gives us a 'rational overlay' on the mathematical and mechanical processes characterized by the theory, an overlay that makes contact with our commonsense conception of ourselves.



Thank you!