Theories of Mental Content, Part I:
Causal Theories of Content

Problem of Intentionality: Explaining how things with intentionality (i.e., aboutness) have this property.

Over the last two weeks, we've looked at a popular strategy for tackling the problem of intentionality:

Thought-First Picture: We can explain the intentionality of language in terms of the intentionality of mental states (intentions, beliefs, and the like).

But this leaves open the question:

Million Dollar Question: What explains the intentionality of mental states? What explains why an individual's thoughts/beliefs/intentions succeed in representing what they represent?

To put it another way: on the thought-first picture, my use of the word "dog" gets to be about dogs because (very roughly) I have intentions to use this word in a way that communicates thoughts/beliefs about dogs. But what explains why my mental concept, DOG, gets to be about dogs (as opposed to cats or nothing at all)?

In the eyes of some philosophers and cognitive scientists, this is one of the deepest and hardest questions about the mind. Over the next two weeks, we'll look at two of the main attempts to answer this:
▪ causal theories (this week)
▪ teleological theories (next week)

But before plunging into causal theories, let's warm up with an initially tempting, but ultimately not very promising approach…

I. Resemblance Theories of Meaning
Close your eyes and form a mental image of your favorite celebrity. The image your form seems to have the property of intentionality – it seems to be about your favorite celebrity. It's initially tempting to think that part of the explanation of this is that your mental image *resembles* your favorite celebrity.

Can we use this insight to solve the problem of problem of intentionality? The idea would be:

Resemblance Theory of Imagistic Representation: An image is about some entity (object, person) x iff the image resembles x.

Now, of course mental images are only one very specific type of mental representation. But some might suggest that other mental representations are implicitly associated with images. For example, perhaps my mental concept DOG is implicitly associated with an image of dogs. If so, maybe these concepts inherit their content from the associated images:

Generalized Resemblance Theory: A mental representation *m* is as about some entity *x* iff *m* is associated with some mental image, and this image resembles *x*.

Despite its initial appeal, the resemblance theory faces a number of problems:

1) Resemblance doesn't seem necessary for representation
   - Is every meaningful concept associated with some mental image? Consider concepts like *justice* or *prime* or *knowledge*. Perhaps these concepts can be associated with images, but need they be?
   - It seems that my mental image of *x* can succeed in being about *x* even if doesn't accurately resemble *x* (perhaps I have mistaken beliefs about what *x* looks like)
2) Resemblance doesn't seem sufficient for representation
   - Resemblance is a *symmetric* relation, whereas representation is not. If my mental image of a dog resembles a dog, then the dog also resembles my mental image. But intuitively the dog isn't about my mental image – indeed, the dog isn't *about* anything!
   - Suppose I form an image of my friend Judy. Unbeknownst to me, Judy has an identical twin Trudy. My mental seems to be about Judy. But doesn't it resemble Trudy just as well?
   - Putnam's example: some extraterrestrials live on a planet that doesn't have trees. These creatures like to make abstract paintings, and it just so happens that their paintings perfectly resemble trees on earth. Do their paintings represent trees?
     o Cf. Putnam on water and XYZ

What then, would be a better theory of representation?

II. From Resemblance Theories to Causal Theories
Think about the last two problems for the resemblance theory. Why is my mental image about Judy, and not her twin Trudy? And why don't the extraterrestrial's paintings represent trees?

Here's a tempting diagnosis:
   - My mental image is about Judy because I have been in causal contact with Judy (I've met her). It's not about Trudy, because I haven't been in causal contact with her (I've never met her, or heard of her, or seen a picture of her).
   - Similarly, the extraterrestrials are causally isolated from trees. Hence their paintings don't represent trees.

Perhaps, then, what's essential for successful representation is some sort of *causal connection*. Here's a simple (too simple, as we'll see shortly) version of this idea:

Simple Causal Theory: A mental representation *m* represents *x* if and only if *x* causes occurrences (aka "tokenings") of *m*.

To illustrate: the idea is that whenever I employ my concept DOG, I am in a particular brain-state. What makes this brain-state about dogs (rather than, say, cats), is that the presence of dogs causes me to be in this brain-state.

The simple causal theory constitutes significant progress over the resemblance theory. (Exercise: go through the various problems for the resemblance theory and work out whether the simple causal theory avoids them.) But it still faces difficulties…

III. Problems for the Simple Causal Theory
1. Many things might cause me to token my concept DOG. For example, foxes or wolves or funny-looking cats my cause me to token this concept. Still, we want to say my concept DOG is about dogs, and not foxes (or wolves, or funny-looking cats).
2. On some occasions, my concept DOG might be caused by a hallucination, or by a meddling neuroscientist tampering with my brain. On these occasions, we don't want to say that my concept DOG is about hallucinated dog-images, or meddling scientists.
3. Following up on 1): if both dogs and foxes cause me to token the concept DOG, then it seems my concept DOG is caused by the disjunctive kind, *dogs-or-foxes*. But presumably we don't want to say that my concept represents this disjunctive kind.
4. When I see a dog, there is a causal pathway through the visual system: I have a certain dog-shaped retinal projection, which in turn causes me token my concept DOG. But we don't want to say that my concept DOG represents the retinal projection. (And we don't want to say that my concept represents the disjunctive kind, *dog-or-retinal-projection-of-a-dog*).
5. On some occasions, my concept DOG might be activated by someone asking a question such as, "What kind of animal is called 'Fido'?" or "What's your favorite animal?" But here too we don't want to say that my concept DOG represents the question (or that it represents the disjunctive kind, *dog-or-dog-related question*).

These difficulties can be lumped into two more general problems:

- *The Problem of Error:* We want to say that my mental representations can be activated in a way that *misrepresents* the environment. But the causal theory doesn't seem to allow for this.
- *The Disjunction Problem*: In a case where my mental representation m is sometimes caused by x and sometimes caused by y, we don't want to be forced to say that my mental representation means *x-or-y*.

IV. Towards a Solution?
For these reasons, the simple causal theory isn't particularly popular. However, some have thought we can mitigate these problems by switching to a more sophisticated causal theory.

One option is to appeal to "normal conditions":

> *Normal Conditions Causal Theory:* A mental representation m represents x if and only if in normal conditions, x will typically cause occurrences of m.

Question: Does the normal conditions causal theory avoid the problems for the simple causal theory? If so, which problems does it avoid? Are there any problems it doesn't solve?