INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

PSYCHOLOGICAL ECOLOGY

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

LEE ANDERSON BASHAM

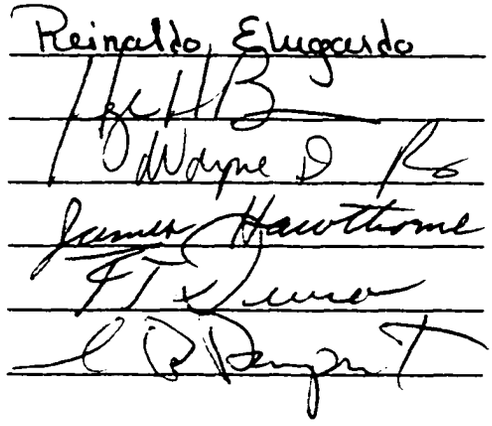
Norman, Oklahoma

1998

PSYCHOLOGICAL ECOLOGY

A Dissertation APPROVED FOR THE
DEPARTMENT OF PHILOSOPHY

BY



ACKNOWLEDGEMENTS

I would like to express my gratitude to my friends and family for their continuing support and encouragement. Especially Roksana Alavi, Mark Eddy, Frank Claesges and Chad Woodruff. I would like to offer special thanks to Mark and Chad for discussion of the some of issues in this work, and Roksana for her love and emotional support. My greatest thanks are reserved for my parents who have steadfastly done whatever they could to make this dissertation possible.

I also wish to express my gratitude to Dr. Reinaldo Elugardo, the chair of my dissertation committee, for his support, confidence, assistance and challenging mind. While holding me to his high standard he also gave me the freedom to make my own mistakes. I would also like to thank the other members of my dissertation committee, Professors Lynn Devenport, Frank Durso, James Hawthorne, Hugh Benson and Wayne Riggs for their support and conversation. In this regard I would like to make special mention of Lynn Devenport, who taught me so much about what a real science of the mind might

look like, and Frank Durso for his patience with a philosophy student in the world of cognitive scientists.

Abstract

A naturalist account of representational content is developed from five related themes. First, an explanation of content should attempt to give those conditions that are relevant and sufficient for the existence of representational content. Second, we should be guided in the resources we appeal to in order to fulfill these conditions by the explanatory opportunity these resources offer. The only serious explanatory opportunity for a generative theory of content is to be found in an account that appeals to a cognitive system's relations to its environment. Third, there are several equally sufficient kinds of environmental relations that appear to generate representational content (causality, indication, resemblance). Each of these can separately, simultaneously act upon the same representation, generating multiple contents for that representation. Different instances of the same content generating environmental relation can also simultaneously act on the same token, again producing multiple contents. This is the core thesis, that the contents of representations are *multiple.* Fourth, there is no principled way to set aside some of these contents. All leading, contemporary naturalist theories of

content fail to do this: Fodor's, Papineau's, Millikan's and Clark's. Fifth, the core problems of any theory of content, the *qua,* depth, disjunction and misrepresentation problems, are easily solved by a view that accepts multiple contents. The conclusion is that a naturalist, generative theory of content has nothing to fear from multiple contents, but is actually advantaged by these.

Table of Contents

Page

Chapter I. Introduction: Representation and Aboutness.

This work develops a theory of *representational content.* It is an exploration of how things like thoughts manage to be *about* anything-- about cats, airplanes, marshmallows and whatever else the world and our imagination holds. There is an ancient and powerful division in our understanding of reality; there is the world, and within us, another world, the world of representation. The following tale draws its surprise and fascination from its denial of this division, its force from how intimately we are connected to *our inner ability to represent,*

God had projected his ideas of "reality" for eons, training the nascent minds whom he had seeded into the void. Then, when he judged the time had arrived, he suddenly withdrew the "world". The divine sensorium vanished--forever. When he removed his training wheels madness erupted from his children. Everything they thought, felt or desired was instantly projected into being, and the massive contradictions within and between them created a howling, raging psychosis. Summer days and solid earth had held quiet sway moments ago. Now terror, rage and delusion stormed. Blinding fear swept up billions of minds into a mind-shattering scream of confusion. The world became a mental meat grinder, crushing the weaker souls within moments. But a handful of strong minds, possessed of extreme clarity of purpose, held tight to a basic ordering and cautiously intertwined their realities with one another (this was a frightfully dangerous thing--a few misthoughts and the cascade of chaos and horror would re-erupt). Together they began to build. Tense but confident, they

entered the state that God intended. They were truly in His image now, creator-offspring of their own new world. Childhood was over.

The key idea here--the ability to represent existence in so many shades and flavors--is our topic. *What is this?* It is hard to imagine a more compelling and fundamental thing to understand about human beings and about oneself.

Within this topic is a complex *workspace,* a collection of fascinating issues critical to how we explain this ability. It's remarkable how complex things can become. But if done well, our understanding will not dissolve into themeless minutiae. Instead we will light up the essential and fundamental points needed for a deep conception of representation's bare existence. In any task one needs a starting point. Mine is "Naturalism" and the question, "How does something exist?". The project is to construct a *Naturalist, generative theory of representational content.* Naturalist? Generative theory? We will detail these ideas shortly. But first we must ask in the most general way, "What is 'representation' and 'content'?"

Representation and content

The concept of a representation is traditionally understood as involving two things: A thing that represents and

what it represents or "stands for". The scheme is metaphorical: Representational government is a government that uses some individuals to "stand for" others in its deliberations. They stood up and spoke for others not present, before the deliberating assembly. S to "stand for" is to act in another's place. Within our minds there are entities that *stand for* things they are not. My thought of a cat isn't a cat, but it stands for it.

In our ordinary speech, our representations are ideas, thoughts, beliefs, perceptions and desires. These are said to have *representational contents*. In some strange sense they "contain" what they stand for. The mind is a vast assembly of representations. By knowing these, combining these, changing these, affirming and rejecting these, we come to understand and control our environment. More grandly:*By our representations we become intimate and master of ourselves and our world.* If we survive in this world, if we understand it and respond intelligently, we do so almost entirely through our ability to represent it and understand the implications of our representations.

Content enjoyers and content bearers ("containers")

To understand representation there are three questions we can explore. "What things have representations?". That is,

what things enclose representations? "What is it in these things that *are* the representations?". What do they enclose? "How do these things manage to be representations?". How is it that these things have the representational power they do?

As for the first question, I will assume anything can contain representations until reasons are given why it can't. Such generosity is a very contemporary attitude, one reflected in our time's enthusiasm for Artificial Intelligence, the consciousness of nonhuman animals and the so far fictional but very serious possibility that radically different *extraterrestrial* entities might have minds. The second question is traditionally answered, "thoughts, desires, feelings and other familiar *mental states.*" Contemporary approaches are often generous on this score too, extending the traditional list to include neurological features, computational structures like those found within your pet computer and input/output functions which don't fall under the traditional mental states. We will not closely examine these expressions of generosity because the ultimate arbiter of their correctness is our third question, *how* it is that something manages to represent. An understanding of this will determine the rest. It will be our focus.

The first step, peeling away the metaphors

How do representations *stand for* anything? This is the essential starting point for any enquiry into representation. Manning our armchairs helps show how natural a starting point this is. Here's a simple picture of an initial reflection on representation,

1) *There is a representation, call it R. R is the occurrence of a concept, idea, etc.. Say, the idea of cats. It is a "representational token", a particular instance of a representation.*[1]

Right. But what makes R *a representation?* As opposed to a Super Nova? What am I saying about R?

2) *R is a representation of cats because it stands for cats.*

Of course it doesn't literally "stand"--again, this is a metaphor funded by the old practice of standing up when speaking in front of a decision-making group. Instead R represents because it has a *content.*

[1] The type/token distinction is at work here: A token is not the type R, but only an example of this type, the way there is this or that particular apple, as opposed to the type of fruit, *apple.*

3) *R represents cats because it has the content,* ***cats***.[2]

So R contains cats. But does this mean I have cats in my head? Very small ones perhaps? No. R doesn't involve a bunch of little cats running around. Our brain surgeons have carefully checked. "Content" is also revealed to be a metaphor. What do we mean by "R has the content 'cats'."? We mean R is about cats. So we turn to *aboutness,*

4) *R's content is what R is about:* ***Cats****. This is how the "stand for" ability manifests itself.*

But what is "aboutness"?

Our answer is a central piece of self-understanding. The idea of *aboutness* holds our image of *the thinking being* together. Aboutness is the universal foundation of cognition. To have a thought is to have something that is *about.* A thought that has literally nothing that it is about (not even the concept of "nothingness") is no thought at all. The same goes for representations like a belief, a memory, a recording, a picture, a dream. Try to imagine any of these without it being about something, be this specific or vague, clear or confused,

[2] In what follows I will mark contents by **bolding** the English words or expressions for them.

consistent or contradictory, possible or impossible, known or unknown, concrete or abstract. We can't give an example without, in the very act, indicating something or other that it is about. *Aboutness is the test for representation. Every representation is about something.*

This is true even if we take a contemporary, minimalist approach that views cognition as simply a species of *computation*. Computation involves variables and their quantifiable relations. Computations are about at least these quantified relations, if nothing else. $F(x) \geq (x-e)^2$ is about the function F's quantified relationship to e.

A. The marks of content and content realism

Marks of content, the syndrome of aboutness

Universal to representation as aboutness is, it is not a clearly defined property. As with so many key ideas in our lives, a pre-philosophic question like, "what is aboutness?" has no ready answer. Instead we find aboutness revealed by three general *marks*.

The first is operational:

Things with content can be used by a cognitive system to act

upon the world in reliable ways--they allow the system to control the world beyond itself in a way that adapts itself to the facts of the world; past, present and most interestingly, future.

The second, related mark is intrinsic:

Content can be (and almost always is) in some sense separate or independent from the thing it represents. First, content has an aspect of autonomy from reality--it can be mistaken or fictitious , involve what no longer or doesn't yet exist, but still be contentful, because it involves something beyond itself, and this something may or may not exist in the world. Second, content manages (paradoxically) to be what it is not: My thought about Little Cat B is not itself a cat. But I can consider and imagine Little Cat B by attending to my thoughts about him. Content is also most always separate from the representation that has it--this dissertation is about content, not about dissertations, the sentence, "Cats are gods." is about cats, not sentences.

The third is experiential:

Content is something we appear to encounter in the way we experience our lives and relations to the world. We

experience things about our world. We apparently have perceptions, thoughts and hopes. They are about our world and lives. We experience learning what things are and how things work through our representations of them. We experience using representations to make contact with reality.[3]

The first "mark" is a general, very beneficial property of representation--it let's us act in accordance with things not directly, immediately, causally connected to us. It gives us a little elbow room in the world. This is especially interesting when we *predict* or *anticipate*. Here the present succeeds in giving us access to the future even though, paradoxically, the future doesn't yet exist. A very nice and useful trick.

The second appears to be *how* representation does this. All too metaphorical in expression, but representation seems to "reach out beyond" the thing that is the representation, to the thing that is represented, all the while remaining distinct from that too. It is a sort of freelance ambassador to the world and whatever else might be, to what isn't existent, and even to the impossible. The "in some sense" is there to allow a great deal of flexibility in how we understand this "separation". Perhaps

[3] Notice that this is a very different, far more impressionistic and far less ambition claim than the view that we somehow directly know in the very act of our cognition the exact representational contents involved in that cognition. I don't believe we have anything like that ability. "What precisely am I thinking?" is not oxymoronic.

everything represents itself, but most contents are more ambitious, and have as their content other things besides the content, **this content**. The ability to reach beyond itself is a *very* interesting property. This ability to be separate from what is represented and separate from the representation could only be had by a *relation*. Aboutness is a *via media*, a way between the thing represented and the representational token. Content and its power to "reach" is a relationship, not a thing in itself.

The third trait has to do with our self-representation--we represent ourselves--apparently at least--as *representers*.We encounter our representational ability in some way that let's us be aware of its existence. What kind of awareness is this? Does it constitute *certain* knowledge? This introduces the important question of anti-realism about content.

Some thoughts about these three: They are not meant to be anything as exact--or grandiose--as necessary and sufficient conditions. Instead they outline a *general syndrome*, a pattern of being and action. The marks are not controversial.*Something like this appears to be going on, something seems to be doing these sorts of things*. Everyone concedes this much. This is what we all know and share, and there is little to add to it. We need a noncontroversial ground to build on. I believe the marks are sufficient for this. To move

beyond these marks is to begin to invoke, how ever vaguely, specific theories of content and to take stands on various issues that are not manifest in our everyday contact with the representational nature of our lives.

Reference, description and truth

Aboutness has a special place in language, where we see the phenomenon explicitly divided into three features: Reference, description and truth-value. Reference maps single representations onto single entities (perhaps hypothetical ones) without providing, via the representation, any information about the entities' properties. Description provides such information, allowing us to understand *what something is like.* Truth-value is a property that, on my preferred understanding (a correspondence view), relates the facts of the entity to the contents of the representation, in terms of whether these contents *match* the facts.[4]

In one venerable tradition the "meaning" of sentences (linguistic sentences) is given by *truth conditions.* What makes a sentence true is what it means. There is a lot of intuitiveness *and* unclarity in this slogan, but putting both aside, I do not believe it is germane to determining how representation exists.

[4] The issues involved in these basic notions are immensely complex. Consider Putnam's views on reference and his resultant theory of "Internal Realism" (discussed in his *Representation and Reality*). Due to topicality and space limitations, I cannot address even a small fraction of all these complexities in this work.

Basic representations are the constituents of sentences; "dog", "ran", "cat" and so on. Pretend (it is a very easy thing) that there are mental representations like these words. Being simpler than sentences they nevertheless represent. Because they needn't necessarily assert or deny anything, they are not the sort of things that can be understood with truth conditions.[5] It may be that a theory of representational content will help us understand truth conditions, but it doesn't start there. Different concerns will drive the enterprise.

An understanding of content should gesture towards an explanation for at least reference and description. The nature of truth is a related, but very involved subject of its own. It would over burden us to demand an account of this too. Suffice to say that truth has only one foot in the realm of content. Understanding how it places its other foot in the *world* is another task.

Skepticism about the existence of representational content

Imagine the denial of the existence of representational content: *Content anti-realism.* A radical proposition but one that must be faced. There are bold and subtle declarations of

[5] One might argue that the meaning of words is determined by their role in determining the truth condition of a sentence. But this clearly gets the process backwards, since words' content are the only thing available to determine the truth conditions of a sentence. They and their contents are prior.

antirealism. The presentation of *bold* anti-realism one might expect is two fold,

> (1) Content does not exist. Nothing has any representational significance at all. (2) Instead content is an illusion of human neurobiology and/or our present conceptual development/cultural indoctrination about human nature.

(1) denies content, (2) gestures towards why we are so misguided; we are in the grips of an illusion. Perhaps this "illusion" can be best made out in this way: Behaviorism, which in its ontological version denied all inner mental states, understood the illusion of mentality in terms of interpretation--mental terms really referred, but only to constellations of bodily movements and histories of stimulation, not to mysterious inner entities. Adapting this to content anti-realism, our illusion may simply be a fact of our neural and speech patterns. Our language systems process neural activity in ways that produce the pattern of utterance that we mistake for a deeper reality, content. These patterns are highly pragmatic yet utterly instrumental. They are not themselves contentful, nor is anything. What we took as representation is actually only a pattern of congruencies between neural patterns, behavior (especially language output) and the external world.

The obvious objection to this kind of skepticism is that its

very expression appears to presuppose content. How can one represent the condition that representation doesn't exist without demonstrating that it does? A proposition is being asserted: The proposition that content exists is false. But propositions, assertions, and falsity all involve content. We are, in effect, being told that telling doesn't exist.

Anti-realists may complain that this begs the question--isn't this "skeptical expression" one of the things the skeptic claims an alternative explanation exists for? I have some sympathy for this move. Nevertheless it is true that the *best explanation* for what is occurring when a person utters, "There is no representational content" is that they are trying to represent in thought and words the nonexistence of representation. Here we have a powerful and proven explanatory means, with great predictive power and a soothing hand the touch of which *rationalizes* much of the behavioral world. And this easily carries the representational *theory* of mind forward.

There is a more decisive point against a content skeptic though. The point of gesturing towards an alternative explanation for the syndrome of representation is critical. Content anti-realism needs to be more than an otherwise mute denial. There is a *syndrome* being got at by content ascriptions.

Realism about content isn't like believing in the Great Pumpkin.There is *something* fairly involved we find exposed in the marks of content, something we have, we're told, made a mistake about. This is exactly what's wrong with the remarks above that describe the "illusion of content" as "only a pattern of congruencies between neural patterns, behavior (especially language output) and the external world.". How the syndrome appears out of all this is unaddressed. So we have two questions; what is *really* going on when we appear to perceive, understand and predict, etc., that we explain with content, and where did we go wrong in developing a vision of this that was based on the idea of content? In the face of the explanatory power of contents, these questions are unavoidable; if content antirealism asks we brush them aside we're more likely to brush it aside.

But then the problem for the anti-realist is that any alternative explanation for all that we have "mistaken" as content needs to show that it isn't in fact an *explanation* of what representational content is, because an explanation of content, if true, would entail the existence of content. The challenge to the skeptic is simple: What is it about content that is a *necessary* condition for its existence and, sadly enough, *fails to exist?* This is what needs to be specified, and any attempt here looks futile.

This is because the representational motif is (1)*so totally basic to our ability to understand ourselves and what we do (we aren't going to easily dismiss it) and* (2)*the fact that beyond a general syndrome we have little prior handle on what exactly this capacity involves (there is no set list of precise necessary conditions for its existence).* Because of this anything able to account for the marks of content will count as an explanation, not an elimination. Recall the marks of content above: There is an ability to in some sense model the past, present and future. There is an ability to be in some sense slightly or radically accurate or inaccurate. There is an experience of undergoing these states. We're not going to give these up. Nor are we necessarily committed to anything much more exact. The upshot is inescapable: *Something stands behind these three "marks of content"; whatever it is, that thing is what we call "representational content".* The would-be skeptic is faced with (1) a target the existence of which we are fundamentally committed to and more importantly, (2) a target that because it lacks definitive, necessary conditions for existence, cannot be pinpointed for a knock-down blow. This second point, (2), is decisive. So we must return to an explanation of the enigmatic power that we call "content".

Subtle anti-realism

A more subtle anti-realism can be developed from Daniel Dennett's notion of an *intentional stance.*[6] Dennett's subject is *belief/desire* explanations for behavior. These are representations with content (they are "intentional states"). Putting aside any questions surrounding the existence of the types of representational tokens called "beliefs and desires", let's just focus on the *contents* involved. Then the view is that content is a mode of description, a way of portrayal that allows unique and powerful kinds of prediction and explanation. It is an *instrument* of explanation and control, and so is extremely valuable, even *unreplaceable by anything else.* But notice that it is possible to continue, claiming that none of this shows that content is a *real, objective, description-independent fact.* Is *Dennett* an anti-realist? Dennett's statements on this issue appear inconsistent. I believe his position is best understood as antirealist, but there's no need to pin this down.The point is the intentional stance funds a subtle and powerful kind of anti-realism if we want it to.

The response here is no different than in the case of *bold* anti-realism. It is true, even if we don't attempt a generative explanation of the marks of content that give the representation

[6] See Dennett's *The Intentional Stance* (Cambridge: MIT Press, 1987), Chapter 2.

such surprising and useful powers, we can still adopt the intentional stance and enjoy its benefits. We can even speculate that there is no reality to this explanatory wonder. But if we go farther, and succeed in correctly explaining *how it is that the marks of content exist (and are not merely described)*, we will be left with a generative theory of content, a realist account of what content is and where it comes from, beyond our descriptions of contents and our uses for these descriptions. This is exactly what we aim to do in the course of this work.

B. Explaining Content: The Possible Projects

What counts as a theory of content? What are we trying to do when we "explain" content?

An initial distinction that is important is that of *derivative and nonderivative content.* Derivative content is content that is derived from an already contentful system. An obvious example would be the coining of a new word or the introduction of a special symbol, like π. Derivative content is parasitic on content already generated by a cognitive system. Our interest is in *nonderivative* content. This content is the sort that we must explain. It is initial.

There are different possible projects here: The definitional, the epistemological and the ontological/generative.

Let's look at each.

The definitional project

Taking a lead from a program in the theory of knowledge, we might believe that a sensible theory of content would determine *necessary and sufficient conditions for each representational type or "concept".* The general formula would look like this,

Representation R has content **C** iff _____ .

where the blank is filled with a either (1) a set of conditions necessary and sufficient for any concept **C**'s content or (2) a set of necessary and sufficient conditions for the *occurrence* of a concept with the content **C**.This latter project was at one time endorsed by Fodor.[7]

Let's examine (1) first. This is a misguided project for the simple reason that it is virtually impossible that most of our concepts are exact enough to support a stringent set of necessary and sufficient conditions. There is no way, beyond blunt stipulation, to give necessary and sufficient conditions for

[7] Fodor explores this project at length in his *Psychosemantics: The Problem of Meaning in Philosophy of Mind* (Cambridge: MIT Press, 1987). See also *A Theory of Content and Other Essays* (Cambridge: MIT Press, 1990), p. 32.

notions like **cat**, **leadership**, **beauty** or most notoriously, **knowledge that *p***. The more general and abstract a concept (barring the stipulation we find in formal logic), the more obvious it becomes that whatever the proper analysis of concepts, necessary and sufficient conditions are totally inadequate.

This needn't be a sign of seriously confused conception or worse, meaninglessness. The task of producing necessary and sufficient conditions is much like the chemist's task of building up atoms from their sub-atomic components and molecules from atoms (the principle of *auf bau* (building up)). It is the task of identifying certain links to other concepts that are at least perfectly coextensive with the concept in question.These "links" and the other concepts involved are always rendered in natural language, and must themselves be decomposed into their components. Is it reasonable to expect this kind of exhaustive, all encompassing *engineering* in fully adequate, meaning-rich natural language? It would be very surprising.

A more likely scenario: Imagine two natural languages, A and B, and reasoners in each trying to determine the necessary and sufficient conditions for x being a *square*. In A the concept of "equal angular" is expressed, but in B it is not. So an endless

variety of slanted figures confront would-be necessary and sufficient conditions in language B as counter-examples, *for it is intuitively obvious that these are not true squares*. Where far more elusive and involved notions are concerned, any language is almost certainly in the position of B.This is even more obvious when we realize that a full analysis in terms of necessary and sufficient conditions would have to apply to the terms used to express the conditions too. The links would have to go all the way down to *absolutely basic* notions. But natural languages are not systems that have any need for a sort of formalistic perfection where everything has some simpler set of terms it can be dissolved into. Languages develop to function and control the world as we find it and live in it--a lexicon where terms are all dissolved into necessary and sufficient conditions is far from what's required to do this. My child's understanding of *come here now!* needn't rest on his grasp of necessary and sufficient conditions for the concept of approach and position in space and time. It is hard to imagine what in ordinary life and our discourse about it would.

All this is supported by a well developed and experimentally confirmed alternative view of concepts, the

prototype theory.[8] Prototype theory sees concepts as *degree of fit gradients* centered on a prototype. Is x a case of **C**? On this view the answer comes in degrees. Is x just like, somewhat like or very little like the prototype of **C**?The prototype is the perfect fit, and the concept shades out from this in terms of degrees of resemblance. The prototype is not itself a set of necessary and sufficient conditions, but a set of variables and their values *that can be tended towards.*

Whatever we make of prototype theory, the quest for necessary and sufficient conditions is at least flawed in supposing a kind of utopian *auf bau* language exists in the strange entities we call English, French, Farsi or Mandarin. There is just no reason to expect necessary and sufficient conditions for R having content **cat** (etc.) are waiting to be unearthed by the persevering armchair pilot.

The concern with interpretation (2) is similar. The attempt to find necessary conditions is over ambitious if we intend to proceed with at least some empirical constraints. The fact may be that nature has multiple, radically independent ways to build content, ways that share with other paths to content few or none of the properties that underlie content's occurrence. We should

[8] For a succinct exposition of this view see Stephen Laurence and Stephen Stich, "Intentionality and Naturalism", reprinted in *Deconstructing the Mind*, (Oxford: Oxford University Press, 1996), pp.168-191.

not prejudge this issue with the definition of our project. Only the pursuit of the project can provide any insight here, not the initial character of the project. There is also the ambiguity in "necessary"--do we mean "necessary given a wholly natural world" (physical necessity), or "logical necessity", or some other metaphysical variant of necessity? An empirical study should be able to proceed without an issue like this being resolved. There is really no theory building guidance in this sort of characterization of our project.

We should also notice that finding conditions that are merely sufficient for the generation of content can be an important achievement in its own right. We can do very well by understanding merely sufficient conditions. Discovering that pressure reduction is sufficient for the precipitation of water from air is an important piece of scientific knowledge, even if it is not knowledge of necessary and sufficient conditions (there are other conditions that can produce the same effect without changing air pressure). Sufficient conditions are going to be the initial harvest from an empirical investigation in virtually any field. Even if we cannot determine if among these conditions are also necessary conditions, we should not therefore reject the account as "inadequate" or "irrelevant". The identification of sufficient conditions may be the best we can ever do given our

epistemic situation. Moral: Necessary conditions are not necessary for good theory.

Consistent with this suggestion, Fodor has more recently offered a watered down version of his original demand necessary and sufficient conditions for the occurrence of a particular content.[9] He agrees that instead all we need is *sufficient conditions for the occurrence of a content,* not necessary ones too.

But a basic criticism of this approach is telling. As necessary and sufficient conditions seem too hard to find, mere sufficient conditions are *too easy.* We can imagine a list of all individuals that are intentional systems at a given time, stick "or" between each member, and then continue the list for all "near" possible worlds (nomologically possible ones, for instance) and finish up with the conditional and consequence, "x has content state y".

This monster will even support counterfactuals. But at best it's just saying: Throw all the physics that goes into creatures with contentful states, throw all the physics that goes into their worlds, and so on for everyone else in all the other physically possible worlds, and you've explained content. I suppose that's reassuring. But empty as an account--"trivial" is

[9] Jerry Fodor, *A Theory of Content and Other Essays.* Cambridge: MIT Press, 1990. See chapter 1.

one word. But what exactly is wrong with this "explanation"? There is no discrimination between facts that are and are not relevant to the existence of any particular content. What we need to know is what, out of this avalanche of facts, is relevant to x having content **C**.

The ontological/generative project

This project seeks to understand just what sort of entity content is and how it is generated in the world. "Ontology" here means what sort of thing content is. "Generation" here means how, from nonrepresentational, noncontentful facts content *is created* in the world. This is our project. In the following chapters we will try to produce an ontological, generative explanation of representational content. This project gives *sufficient conditions of a very special kind, the kind that actually constitutes the existence of* ***C***. This is a demand for sufficiency tempered by relevance and what I intend by the expression "a generative explanation".

There are subtleties here. A generative explanation is easily confused with a causal-historical one. The distinction is similar to the distinction between "how it got here" as opposed to "what is it and how does it endure". Whether our interest is in objects, or *events, activities or relational properties* we will see

generative explanations as akin to compositional scenarios. These scenarios will include objects, their properties and the relationships these enter into.

Take carbon for example. A causal-historical explanation for carbon assumes its generative explanation and tells us how this situation came about: Gravity driven nuclear fusion inside stars produced the carbon in my body. But what is the ordering of facts that makes carbon what it is, moment to moment? What is the generative explanation for the fact that the carbon in my hand exists now *as carbon*, and *endures into the next nanosecond as such?* A causal-historical explanation for carbon's presence in my body cannot answer this. Beyond knowing "where carbon comes from" we can ask what entities and their relations *constitute the existence of carbon.* The answer is that carbon is the atom constituted by the following subatomic objects x, y and z and their interactions. Whether our interest is in *object,event, activity or relational property*, we will see generative explanations are powerful sources of understanding and much of science is bent upon their discovery.

The aim of a generative explanation for content

What then does a ontological/generative explanation of

content aim to explain? *It aims to explain the origin of the syndrome of aboutness. This is critical when we try to evaluate just what might be contenders for a generative source for content. Given what we know prior to our theorizing, this is the only available target of explanation. We are simply trying to explain the syndrome. All else is post -theoretic.*

The distinction between the ontological and generative and the "just does" answer to naturalist representation.

The ontological and generative are *importantly* distinct questions for the simple reason that content might have a nature but no interesting generative facts about it; it might be inexplicable in generative terms, and instead be a *basic* fact about reality: Content "just is" part of the natural order and there is no generative explanation for this in terms of nonrepresentational facts.Everyone who is a content realist must admit the ontological question is a powerfully important one, especially for metaphysics. But the generational question is more controversial.

We can trace this controversy back to the great British philosopher Thomas Reid.[10] Drawing on Reid's writings, in a recent article Michael Tye has defended the nongenerative

[10] See Thomas Reid's *Inquiries and Essays*, (eds.) R. E. Beanblossom and K. Lehrer (Indianapolis: Hackett, 1983), *An Inquiry:* chapter 5 *(Of Touch)*, section III *(of Natural Signs)*, pp.43-44.

approach with vision and intelligence. Tye puts it succinctly,

"....what we should believe....is that our capacity to think about things....is a natural phenomenon whose operation we cannot further explain by conceptual reflection."[11]

After rejecting the definitional projects for reasons very much like the ones explored, Tye gives a simple argument for the failure of *any* explanation of content. He takes neurological accounts and functionalist theories of mind as the contenders to meet. He rejects both with a simple argument from *multiple instantiation.* Most famously deployed against the *identity theory* of mind, the argument is this,[12]

P1) Any identity or Functionalist theory of mind (call it "T") claims that *all* mental states/processes are identical to a single sort of entity, or a restricted class of entities.
P2) Mental states are virtually unlimited in the ways that they can be naturally produced ("instantiated"). Some of these are not the single sort of entity or restricted class of entities mentioned in T.
C1) So there are some entities that produce (instantiate) mental

[11] Michael Tye, "Naturalism and the Mental", *Mind*, Vol. 101. (1992): 421-441.
[12] Ibid., pp 425-426.

states that are not included in theory T.

P3) Any theory that entails a false consequence is false.

C2) All identity and functionalist theories (all versions of T) are false.

The key premise is (2). The idea is that we can always imagine creatures that have a certain mental state (joy, fear, seeing *blue*), but that are either not neurophysiological constructed the way our identity theory says they should be, or differ in their functionalist characterization from what our theories say they should be. One can easily question why this point is sufficient for premise 2, though. We need another premise, something like, all the things (or at least contentful systems) we imagine really are possible. This is doubtful. Imagining a scenario here is not to imagine systems in a way that thoroughly specifies their architecture and inner dynamic (at least as thoroughly as necessary to infer their representational capacities) and shows that these systems have contents that our favorite theory denies them. Instead, we are simply "imagining"--stipulating without *obvious* incoherence--that it is true that, "there are systems that have contents any identity or functionalist theory says they don't." But is it true? We need to know a whole lot more about the particular system and ask, "Are the properties of system x

sufficient to have content y that theory T says the system doesn't have?" *That's the question.* Tye doesn't say. We are certainly not in any position to do so. The problem is that we have no *direct* knowledge of a system's contents. All of our understanding, speculation, reasonable conclusions and so on are based on inferences driven by what we can find out about the nonrepresentational facts of the system--probably facts about neurology, computational profiles, ecological situation, behavioral outputs and so on.

This makes things difficult. How do we know x has a certain content? Behavior is the initial means. Once familiarity brings the system to our attention, it drives us to ask whether the system has contents. Now Tye can't say that behavior, or anything else nonrepresentational in nature, is *constituent* of content. That would constitute a *nonbasic* approach to content. Even if he said different things about different systems without commenting on a unifying commonality between them all, our generative theory would be a disjunction of all the individual accounts. So he must say that behavior *shows* (in the sense of good evidence, not certainty) that the system is contentful. But how exactly? We still must ask--does the system *merely mimic* having content or does it have real content? We need to show with some decisiveness that such and such system actually has

such and such content. Behavior alone is notoriously incapable of doing this. To do this we have to go inside the system, and perhaps outside too--into its environmental /ecological setting--and demonstrate how these facts are insufficient to generate a certain content. To do this we must have exactly what Tye says we can't--a generative theory of content in some degree of development--in order to determine whether an unanalyzed but possible system has or fails to have the said contents. Tye can't make sense out of the process of distinguishing contentful systems from those which don't possess the contents they seem to.

Finally, the universalness that Tye demands in (P1) is suspect. Suppose I have a fine theory of content generation for birds and mammals. To my surprise, but not disappointment (job security) I discover my excellent theory doesn't extend to Martians, digital devices or Republicans (which are malevolent aliens incognito, after all). Does this mean that I should abandon my mammalian/avian theory? No, it may simply reflect that the nature of content might be *domain relative*. This is hardly a first in good science. (An understanding of human reproduction extends poorly to the question of how lichens multiply.)There may even be deep reasons for this that we will also discover. In the meantime, we will develop generative

theories within domains and extend these as far as we can.

It is equally important to realize this doesn't show that we *can* have a generative theory of content. It just shows Tye's argument doesn't prove we can't. Maybe his view is best taken as a sort of open-ended skepticism, without any proof but still worrisome in its possibility. This view is reasonable enough--nay saying is valuable as a caution and as a spur in one's side--but it has two serious weaknesses.

The first is that it is an inherent looser to any plausible generative theory of content that does come along--even a *false one*. If you say content "just is" and I say, "No, here is how it is generated out of non-contentful facts" and my solution is reasonable and satisfactory, we will automatically reject the "just is" claim. This puts the "just is" position on perpetually shaky ground. To escape this "shaky ground" one must show that no generative explanation is *possible*. And this is not a task likely to succeed, so little do we really understand about representation *prior to exploring and explaining its generation.* In the following chapters we will explore a generative explanation for content that I believe is satisfactory.

The second sort of weakness we will explore in detail later. But a quick indication is to the point now: Our representational contents and the world we find ourselves in

are in remarkably good agreement. The door is right there where I represent it, the cars passing by are, yes, actually cars, and my cat is really and truly puncturing my leg again....When we ask *how it is* that my representation's existence and pattern of activation so well *matches* the world, the "just is" view cannot respond.

For instance, I have learned what a cat is. This means I have at least a certain neural structure that has been developed and modified in detailed ways, and which is the token that has the content ***cat***. My representation becomes finer and finer in its discriminations, I no longer confuse cats with dogs and rats. In concert with environmental relations, my brain must be operating along certain principles of neural structure refinement that lead it to a stable state with the content **cat**. These principles relate nonrepresentational, noncontentful aspects of the world--environmental input and neural structure alteration--to the development of content. I imagine they can even be described in some present or future scientific taxonomy. But on the "just is" view, there can be no such principles, because appeal to these principles would constitute a generative explanation of why such and such structure had the content it did. Instead, the learning brain must be *good at guessing*. Multiply this point across every domain of content learning and

refinement:*Here the odds against the "just is" view are overwhelming, so much so that it looks absurd.*

Simple points like this show we really need a generative answer to the nature of representational content. Otherwise we look like addicts to a vast and improbable benevolence in the universe, something like the divinely guided *parallelism* of the eighteenth century.

The epistemological project

The epistemological project explores how we can *determine and verbally express* the content of R . Ideally, this issue ultimately is addressed by the ontological/generative project. But in practice it has proven the favorite foil of skeptics about our ability to specify contents.

Suppose we believed that there was no way to determine what the content of a representation is with the sort of evidence we have access to. Then a generative explanation of content could never be made, at least not in a way we could have any epistemic confidence in. The basic claim, at least the one that is relevant for our purposes, is that we cannot know what the content of a representational state or process is in a way that is acceptable for the purposes of a reasonable, scientific psychology. I will not discuss the often complex and interesting

views here in detail, just comment on the importance of the general project.

Stich is well known for arguing that it is far from clear what any animal believes.[13] This is the kind of case he has in mind: If a dog chases a cat up a tree, does the dog think **the cat has taken shelter in a tall green thing**, or that **the hisser-scratcher-tree-climber has climbed up the tree** or the **yummy furry wants to be dined on up high** (etc.)? Stich claims that there is really no obvious reason to ascribe one or the other beliefs, even though they differ in content.

Quine is famous for his claims about the *indeterminacy of translation between natural languages.*[14] He imagines a cultural anthropologist studying a tribe that refers to rabbits. Their utterance for rabbit has many possible interpretations; rabbit, rabbit time slice, collection of undetached rabbit parts....which is what they actually mean by rabbit? He claims nothing in their behavior can reveal the answer. His famous illustration of a *translation manual* comes in at this point: We can imagine two translation manuals, each of which fits the behavioral evidence equally well, and each of which translates the expressions of the tribe into English in substantially different ways. He infers

[13] Stephen Stich, "Do Animals Have Beliefs?", *Australian Journal of Philosophy*, Vol. 51, (1979).

[14] W. V. O. Quine, *Word and Object*, Cambridge: MIT Press, 1960, especially chapter 2. An excellent discussion of the complexities in Quine can be found in William Lycan's *Logical Form in Natural Language* (Cambridge: MIT Press, 1986), Chapter 9.

(with additional premises, including a commitment to a behaviorist analysis of mentality) that there is no fact of the matter what the tribe actually means.[15] Again, his conclusions can be debated. But let's grant the point, admitting that it is at least hideously difficult.

This skeptical project, while important, is inadequate to its own task. This is because we really can't judge the prospects for determining content until we have a fairly well developed *ontological and generational theory* of content to ask our questions from within. It is always relative to particular ontological/generative theories that our epistemology must encounter its power and limits. It is one thing to ask, prior to such theory, "how can we determine the content of R?". We explore, in effect, a series of half-baked ideas about what content might be and how it might arise, and comment on our access to these in an open ended, speculative manner. It is a very different question when we ask, "given such and such well developed views--say current cognitive science and linguistics and a particular philosophical framework--how do we fix a particular content given our experimental and other evidence?". The ontological and epistemological questions are actually deeply interdependent and deserve equal attention, co-

[15] Ibid., chapter 2.

evolving one with the other as research continues. We will certainly see this strategy when we develop a generative explanation of content in the following chapters.

But this is all a methodological point. In terms of facts, *ontological* facts will always determine the *epistemological* ones, not vise versa. What exists in the broadest sense, including not just what content is and how it appears but also what we have at our disposal in terms of reasoning, perception and the fortunes of theoretical starting place, is what determines the things we can reasonably infer.

In fairness, Quine seems to appreciate this via his commitment to behaviorism, but his philosophical version is primitive--his imaginary anthropologist is sitting on a tree stump with a note pad. He doesn't really engage the advanced, well controlled experimental procedures that any behaviorist psychologist would want to deploy, ideally dissecting the entire stimulus history and reaction profiles of the tribe and comparing them to our own. There is really no telling what an army of caffeinated experimental psychologists might come up with to both distinguish and *fix* the content of an expression that means "rabbit time slice" and one that means "rabbit". The behavioral details that are found to map these contents perfectly may be radically unintuitive and seem just plain arbitrary and bizarre.

None of this shows the impossibility of a precise, one to one mapping of behaviorist variables to contents. More relevant to our present situation, Quine's behaviorism also exists in isolation from the real range of our explanatory alternatives; they far outstrip its possibilities. Vast potential resources are available from current neuroscience, cognitive science and behavioral ecology and ethology.

Finally, and *most important,* there is the possibility of accepting that differing translation manuals might exist and be indistinguishable in terms of behavior and even neurological evidence, and yet the possibility remains that far from the meaning of the tribal utterances being indeterminate, it is multiple: *Both* manuals are correct in the most literal way. The price of this move is abandoning the idea that every representational token has a single content. I will not pursue this idea further now. I only suggest that it is a *very interesting one.*

So before we approach the epistemological issue, we need to develop the ontological and generative theory given the background information we have. This background includes the basic metaphysics from which we must construct content, the experiential and behavioral evidence we have of the apparent contents of our thoughts and perceptions, as well as the indications of coherence and simplicity. Then we can ask what

might be known.

Revising our ideas about content; how far dare we go?

How far can we take our picture of content into unexpected forms and claims? How much of our ordinary understanding is beyond challenge? We should be willing to be very bold in our revisions, as long as we don't loose sight of certain essential facts about content. Here's an example of what we *don't* want,

"I have a theory about what contents are. According to my theory all contents (like the proposition P) have physical essences F, G and H. Of course, nothing that has F, G or H can be a bearer of truth value, can be a description or refer to anything. So, since my theory has predictive power about what people are likely to do, it has the presumption of truth in its favor. So, all things being equal, you need to revise your conception of contents as truth bearers, things with reference or descriptive power, and start thinking of them as things with F, G and H. They'll still be contents, just not things with truth values, reference or descriptive power."[16]

This has a ring of incoherence. An ancient Greek asks another for a definition of a circle and it comes out having

[16] I owe this expression of the point to Ray Elugardo.

exactly four sides. Such a theory would be a fascinating achievement, but it doesn't tell us about representation, because aboutness is out of the picture. We've already seen anti-realism is a nonstarter, so such a theory is also irrelevant to our project. Truth, reference and description are the basic forms aboutness appears in. Abandon these forms and we are no longer pursuing content. This especially applies to reference and description. To the extent our minds have referential contents and descriptive contents, our understanding must accommodate this.

C. What Can Represent? The Ecumenical Attitude

There are two senses to the question, "what can represent?" The first is what kinds of things are representations, and the second is what sorts of entities can have and use representations.

On the first question I believe an *ecumenical attitude* is best--I would even say it is the only honest one.[17] Many things might be nonderivative representations. This might occur in any number of ways. Beliefs and desires are the folk-classics, but

[17] Webster's offers this as one definition of "ecumenical": "promoting or tending towards worldwide Christian unity or cooperation". In its historical context, as well as ours, the word connatates *open mindedness, tolerance* and the relaxing of *presupposition*.

vector multiplications, patterns of crystallization on a piece of chemical film, even the twisting of a flame or the particular porousness of sand stone might all bear primary representative powers. How could we *show* they don't?

Notice that a *showing* is a very different and far more demanding thing than stories that elicit the opinion or "intuition" that something doesn't represent. This point is particularly pregnant because it reveals how little we have in the way of direct investigative tools. Descartes appeared to believe we knew nothing so well as thought itself, but he was quite mistaken. We know more about "the body", and have ways of learning about the body, that far exceed anything we can throw at the question of representation. We can weigh, strength test, dissect, genetically explain and evolutionarily understand the body in obvious ways. Thought is far more elusive.

But humble honesty aside, the real virtue of the ecumenical attitude is the way it allows us *the fullest possible explanatory opportunity by freeing us from a study restricted to uncontroversial but horribly complex representational systems--humans and other advanced mammals*. The keys to understanding representation are present in the *simplest examples,* where what is critical and relevant exists unobscured by elaboration, redundancies, convolutions and the other

machinery of distraction. These simplest systems may be far removed from our familiar, complex human experiences of representation, but the ecumenical attitude lets us examine and profit from them. This way we can explore the foundations of representation without having to tackle the monstrous complexity of our own mammalian capacities, right from the beginning.

Consciousness and representation

Some have strenuously resisted the ecumenical attitude. Writers like Barbara Hannan and John Searle have tried to put that elusive hallmark of the "higher animals",*consciousness*, at the center of our exploration.[16] On a simple version of these views, perhaps we know at least this: *Only conscious experiences nonderivatively represent.* This objection to the ecumenical attitude appeals to our ordinary way of encountering and understanding our representations. We appear to encounter representation most intimately in our "consciousness". We all know what representation is--its all that conscious thought, desire, perception and the rest that courses through us. Perhaps conscious presentation is a necessary

[16] See Searle's *Intentionality* (Cambridge: Cambridge University Press, 1983) and his *The Rediscovery of the Mind* (Cambridge: MIT Press, 1992). The entirety of these works bears on the issue. Barbara Hannan concurs, adding arguments of her own in *Subjectivity and Reduction* (Westview Press, 1994), pp. 81-98.

condition for representation.

There are a number of problems. First is the lack of simple evidence. Our conscious experience doesn't carry the message, "Representations only happen this way." Representations can be conscious. This doesn't in any way imply that they can't be nonconscious. Much more is needed.

The second problem is that there are apparent counter examples drawn from the unconscious; assumptions, desires, memories and so on that are not, even never, conscious but are still representing certain things--what it is that is believed, wanted or has occurred. More technical examples might be the brain stem's monitoring of blood sugar levels, or the immune system's ability to represent pathogen types and store these representations for future use in the production of anti-bodies. One might deny the existence of the unconscious mind, but this goes against the common experience of *discovering* the content of states that were previously unconscious. When we become aware of an unconscious assumption, desire or etc., we *discover* what it is about--what its content is. We don't create the content, thereby creating the assumption. This is only possible if these assumptions and so on had content *prior* to their being made consciously manifest.

Also consider unconscious processing. Because it

involves what looks like the storage and use of "knowledge" or "information", it appears to require unconscious representation. A compelling example is "implicit learning".[19] Imagine an experiment where subjects are presented with pairs of images and perform a simple task--determining whether an image is spherical or octagonal. The time it takes to make this judgment and strike the appropriate button is measured (the *reaction time*).The first image includes a shape and the color of the lines that make up the shape. As it happens, there is a 80% probability that when the color is red, the image after the next image will be a sphere and an 80% probability that when the color is blue, the next will be square. So there is a *predictor* within the previous shapes about the later images. But subjects are not told this. The experiment proceeds by showing subjects images in rapid succession. With the predictors in place the subjects quickly and dramatically improve their reaction times, but when the predictors are removed, these reaction times suddenly plummet to the level found when there is only a random correlation of color to next shape. What's really interesting is that subjects never become aware on their own of the predictive relationship. They are surprised when it is brought to their attention and adamantly deny being aware of it

[19] See, for instance, Dienes & Berry. "Implicit Learning: Below the Subjective Threshold," Psychonomic Bulletin & Review, 1997, 4(1), pp. 3-23.

during the experiment. But clearly it was being used to facilitate their reactions.

How do we understand this? A natural explanation is that the predictive relationship is being unconsciously represented. This representation consists of something like a priming of the associated shape response by the visual cortex registration of a certain color in the present image, based on the fact that the unconscious processing system had abstracted the patterns of correlation from previous presentations.

An abstraction of a correlation and the application of this to judgments certainly looks like representational power at work. This abstraction is in place whether or not we ever notice it and even if, for whatever reason, we *could not* be made to notice it.

A last point, one that I believe is most important; when we look for *positive evidence* that all representational states must be conscious we stumble on the fact that without a serious generative theory of content, there is no way to know this. To show a system cannot generate content we need to show it lacks a necessary condition for generation. Without a well developed understanding of how it is content is generated, we cannot show that certain systems lack this necessary condition. If our necessary condition is consciousness, we need to be

shown *how it is* that consciousness and only consciousness can generate content. We have no such account. Given what little we can state with any certainty about the workings of consciousness, this is an approach that holds little promise beyond an expression of a narrow vision extracted from mere familiarity.

Content and perspectival, first person properties

Searle has tried to respond to this kind if criticism. Again, the basic weakness in tying representation directly to consciousness is that though representations can be conscious, this doesn't imply that they can't be nonconscious. Instead the argument needs to proceed along the lines that some *necessary property* of representation can exist only relative to consciousness. A possible candidate: It appears that our representations often have the property of being given with a certain *perspective* or *"aspectual shape"*. Frege's famous example of the morning and evening star illustrates the idea. The fact is that the star called the "morning star" is identical to the "evening star". Both are the planet Venus. But in representing x as the morning star we do not represent x as the evening star. We had to find out that they were the same object. A similar example used by Searle contrasts the representation

of H_2O with "Water". One can represent water without representing that stuff *as* H_2O, or vise versa. The two representations have different "aspectual shapes". Different aspects of the same entity are invoked in representations that are about the same thing, but have different aspectual shapes.[20]

There are many issues here we don't have space to outline, let alone discuss. What is germane to the role of consciousness is the claim that all contents have an aspectual shape and that only consciousness allows us to account for this. Searle offers this argument,

1) There can be no content without that content having a particular aspectual shape.
2) So all content has aspectual shape as a necessary condition for its existence.
3) Nothing nonconscious has aspectual shape while nonconscious.
4) Aspectual shapes do exist when conscious.
5) So, all contents exist only when provided with an aspectual shape by their actual (or potential) conscious manifestation.[21]

Somehow, it must be consciousness that is doing the

[20] In *Rediscovery of the Mind*, p. 156.
[21] *Ibid.*,156-159.Searle adds "potential" to allow for content ascription to beliefs and desires that are not currently conscious or are never actually manifested in consciousness.

"aspectual trick". This is a deductively valid argument. But is it sound? Let's momentarily grant (1) and (2). (3) is what's really interesting. Why does Searle believe it? After rightly dismissing the idea that aspectual shape is a matter of mere *behavior* (even linguistic behavior) he turns to the brain,

> ".....no amount of neurological facts constitute aspectual facts. Even if we had a perfect science of the brain....all the same there would still be an inference--we would still have to have some law-like connection that would enable us to infer from our observations of the neural architecture and neuron firings that they were realizations of the desire for water and not of the desire for H_2O....the specification of the neurophysiological in neurophysiological terms is not yet a specification of the intentional."[22]

This puts consciousness and its role in producing aspectual shape at center stage. To understand representation we need to embark on a study of consciousness.

Though blessed by Searle's wonderful skill at summoning the intuitive, I believe this argument is ultimately misguided and only creates a needless distraction. Some worries:

First, it is entirely possible that conscious representation needn't involve anything like *aspectual shape.* What exactly is the aspectual shape of a memory of a particular shade of blue?

[22] Ibid., pp. 158-159.

I can just as well recall the shade, and not *as* anything else but what it is, that shade of blue. But it is certainly *about* blue. Representations that Indicate needn't involve aspectual shape either. We can know that x indicates y, and be done with it. X needn't have an aspectual shape, nor indicate y as anything in particular. We might not know anything y on being told that x is indicative of y. But both x and y do represent particular entities. But neither x nor y have any aspectual shape whatsoever. Premise is (1) looks to be false.

But this is nit picking. A deeper point: Even if all *conscious* representation involves aspectual shape, this doesn't imply all representation must. A broader understanding of representation is possible, one that extends beyond consciousness.The *psychological ecology* model outlined in Chapter 2 is just such an understanding. Again, there are obvious doubts about (1).

We can also question the dependence of the "aspectual shape" of contents on consciousness--a dependence supposedly evidenced by premises (3) and (4). We've seen Searle argues for this dependence by claiming that no facts about the behavior or neurophysiology or whatever other nonconscious descriptions we might deploy can, *by themselves*, allow us to determine what the aspectual shape of

a state or process in a system is. He concludes that it is only in the conscious manifestation of these that aspectual shape appear. Let's grant the premise that we can't determine aspectual shape from neurophysiology. Does the conclusion that aspectual shape isn't in fact neurologically constituted follow? It is hard to see how.

Perhaps Searle assumes that if x is present in a body of facts we can tell it is. But this is a *very* exaggerated view of our abilities, so much so that it is plainly false. From the lack of epistemic access we cannot directly infer the nonexistence of something. A few simple observations make a compelling case that the *Earth is round.* But it took uncommon humans and some surprising twists of reasoning to realize the message present in the facts. A highly magnified computer graphic shows us only a swarm of pixels. Even upon seeing the entire image at this magnification we can't tell what it is. Our visual memory and compiling abilities fall far short of what is required. We reduce this information load by standing back, allowing us to discern large scale patterns. Simply because aspectual shape is not manifest to us in a vast, alien and really very intimidating body of neurophysiological minutiae, it doesn't follow that it is not nevertheless present in and constituted by those very facts, independent of any consciousness associated with them.

Searle also puts weight on the fact that an "inference" must be made by us from the neurological to the aspectual. Is this inference what shows that the neurological facts don't constitute aspectual shape? Again, it is hard to see how it would. The principle at work seems to be: If a body of facts (A) cannot yield knowledge of some fact (B) without our *inferring* (B) from (A), then neither (A) nor any part of (A) constitutes (B).

I cannot see how this principle is true. When I infer, via certain scientific background beliefs and explanatory power, the existence of an ancient, now stranded coral reef (B) from a 500m cliff of fossil corals (A), it doesn't follow that the coral reef (B) is not *constituted* of these very corals, this very cliff (A). *My inferential behavior has nothing to do with it.*

So it appears Searle has not given us a good argument for his claim that aspectual shape is dependent on consciousness. But this doesn't show he's mistaken. Can we find reasons to doubt the link between consciousness and aspectual shape? I believe we can.

Nonconscious aspectual shapes

The real issues surrounding consciousness and aspectual shape are these: First, why is it *impossible* for a nonconscious being to represent Venus as **the morning star**,

or a type of liquid as **T-4**, **position 1467a** in the typology matrix, or even **water**, as opposed to **H_2O**? Second, have we good reason to only ascribe content to states and processes by recourse to consciousness? That is, do conscious properties really reveal content in a way that is not equally had by nonconscious properties?

About the first question: Perspectival properties like aspectual shape seem like an *access relationship* to sorts of information or a range of facts. Perspective is a matter of *approach*. Access relationships are not obviously restricted to only conscious systems. In perception this "access relationship" might be realized via causal connections to the environment, ones that produce effects in the cognizer that are keyed to certain particular facts about a liquid and not others. In *unconscious* perception these effects would constitute the unconscious aspectual shape of the perception. In *conscious* perception they would produce the conscious aspectual shape. Indeed, in our own case this is almost surely the situation. But these effects needn't be conscious at all. Now what exactly does consciousness of these effects add to the story? Apparently nothing. The poverty and riches are the same. Electromagnetic and sonic spectrum sensitivities, the mundane matter of placement in a spatial setting, thickness of the skin, lighting in

the room, the neural development needed to draw the requisite discriminations as reflected in neural response, the capacity of a system to retain the ability to make these responses for minutes, weeks or years, all these bless or curse a system without heed to whether it is conscious or not. They may all attach a perspectival mark to the contents. Returning to Searle's argument, premise (3) looks to be false.

About the second question: As far as what we are really able to know about aspectual shape, is attention to consciousness all that illuminating? Searle believes facts about consciousness determine aspectual shape. But what are these? The only ones apparently at work here are the *verbal experiences of inner dialogue*. But these are an extremely weak basis for ascription of aspectual shape. Inner linguistic behavior seems no more privileged than outer, and Searle has, rightfully, rejected the suggestion that linguistic behavior determines aspectual shape.[23]

This point is worth discussing at length. From what I can gather from myself and others, when we "think 'that's water'" (as opposed to 'that's H_2O') this is akin to "hearing a word in your head" while we are viewing something and classifying it by kind. This isn't to deny that one could be undergoing a

[23] Ibid., pp. 157-158.

conscious aspectual shape like the one **water** involves and still be unable to label it as such, or in any way. Various kinds of brain damage can almost certainly produce that effect. The question is how we *consciously know* aspectual shape. The answer appears to be via conscious word assignment or some similar linguistic maneuver.

If this is our data, then determining aspectual shape still eludes our grasp. Suppose I ask Kathy, who is gulping down the fluid that the drinking fountain emits, "How did you represent that stuff, as *H_2O* or as *water* ?". She'll probably say, "water". Fair enough, but what does this come too? We assign words to aspectual shapes, but these words hardly settle what the aspectual shape actually is. If we press her, "How do you know you didn't represent it as what we verbally describe, 'the clear, cool stuff that doesn't taste like much and just quenches my thirst'?" she might reply, "I just thought, 'water', not all that." Suppose she does. We can easily press further, "Isn't what you mean by 'water' the cool, clear etc.?" Here the epistemic problem is explicitly revealed. She might reasonably say either "yes" or "no" to this suggestion. It isn't *obvious* what she should say. Is she merely reporting the word assignment that her language cortex produced? Or does she have conscious access to compelling *reasons* why this is the correct

assignment? How do we determine if this assignment is accurate? That is, what exactly do we mean by an aspectual shape like **water** (or even **H_2O**)? Whatever she says, what facts about her determine if she is right or wrong?

For Searle these would have to be facts about her conscious state at the time she was drinking. But again, just which ones? This flurry of difficult questions illustrates how it is far from clear that an appeal to consciousness gives us ready and definitive tools for determining aspectual shapes. Notice this is true even in our own, first person case.

In this vein, we can imagine even more complicated scenarios. Suppose Kathy was a chemist--until the terrible accident with the hair drier. Now she has brain damage and is not able to consciously access memories of H_2O and other molecular descriptions. But there is some pretty powerful evidence that the memories are still there in some sense; when we ask Kathy to point at various strings of letters and numbers when presented with household items like water, bleach and lighter fluid, she is amazingly accurate--but she experiences and reports this as purely *guessing.* She says these are whims, "in my arms and hands" of which string to point to, "not anything I believe". Brain scans reveal a different story. Whenever she is presented with water, etc., the brain processes that used to

ignite when she saw things as their molecular "aspectual shape" still do, but due to some torn neural links, these no longer play any role in what she is verbally aware of. [24]

Question: Is she representing the fountain fluid as **water**, **H_2O** or as both? She can only report the former, but all the neural activities --short of verbal access to them--are occurring for representing it as H_2O. "Water" is what Kathy reports. She experiences this word assignment and believes it is correct. Most of us would agree this is good (but not certain) reason to believe she represents it as **water**. In contrast, she sincerely denies any knowledge of H_2O as a substance, "Huh? What's that?". It is not available to her as a word assignment. But why not claim the following: She nevertheless does possess the aspectual shape expressed "H_2O" *simply in virtue of these neurological process.*[25] Why should consciously accessible word assignment mark the line between aspectual shape and its absence?

I can't see that it does. Nonverbal animals have no word assignment, but presumably represent. Furthermore, there seems no reason to deny these representations have aspectual shapes. Give chipmunks a pile of sunflower seeds and presumably they represent the seeds as something like *food.*

[24] This is very much like "blind sight" and other cases of neurological access damage.
[25] And external relationships and/or embedding processes.

Dump this same pile onto the entrance of their burrow and presumably they represent the seeds as something like *stuff in the way.* [26] Word assignment doesn't appear to be necessary for aspectual shape.

It is not always sufficient, either. A typical scenario: We experience "what something is"--we feel we recognize it and can say something of its properties. But the designation for it eludes us. Word assignment is fallible, because the mechanism of its production is neurophysiological. Neurophysiology can always misfire. So Kathy might *verbally,* consciously think, "Seltzer" or even "Dirt", even when she actually was consciously representing the fluid as H_2O, only to then correct herself (or conceivable fail to). Is the view then that correct word assignment makes for aspectual shape? But then there are facts besides the verbal, conscious ones which determine correctness. Because there can be no change in the conscious facts without a corresponding change in the nonconscious, neurological facts, these *correctness determining* facts are ultimately present in the nonconscious neural processes. So what does Kathy represent the fluid at the fountain as? Apparently she represents it as *both water and H_2O*, since

[26] I don't believe there is any problem in the fact we can't *say* exactly what the chipmunk's aspectual shapes are. As the points above and below suggest, I don't think we should be all that certain that what we say about our own representation exactly captures the aspectual shapes involved.

water appears to be correctly present in verbal consciousness and the correct assignment to the nonconscious neural processes is **H_2O**.

The key point is that if consciousness is responding to these nonconscious neural activities then there must be some facts about them that determine the correct aspectual shape to adopt. But then there *are* nonconscious facts that are sufficient to fix aspectual shape. To know these we just need to know the conscious system's preconscious, response-scheme (when it responds accurately). I see no reason why this pattern of facts can't be represented in third person ways. If so, then we have a description of an "aspectual shape" that at no point relies on the first person, subjective description or knowledge.

Again, what we need to pin down the answers to all of this is a generative theory of content. Because this is a *generative* theory, it will appeal to the nonrepresentational facts--it will not start with the conscious experience of aspectual shape. It will start with whatever facts determine aspectual shape. There is no prior reason to suspect these facts will be *nonrepresentational states of consciousness*, (qualia like "ouch" and "blue"....). It is hard to see how these could be relevant. Instead a different system of facts will probably be

appealed to, which need include no conscious ones at all. [27]

Conclusion; setting aside consciousness

The conclusion is that we have no compelling reason to burden a generative/ontological theory of representation with an exploration of certain obscure powers of consciousness. Until we are *forced* to widen our explanatory task and wildly expand our theoretical challenge--and Searle's views are far from doing that--I suggest we *don't.*

Other worries about an ecumenical attitude, representation as "good for something" and panpsychism

Some might disagree with the ecumenical attitude for other reasons. They might argue that representation only makes sense in highly flexible creatures, beings that can execute the subtle and complicated behaviors thought makes possible and so can benefit from the immense neural growth and maintenance expense it takes to field a cognitive system and keep it running.

This would be a fine point if we replaced "representation" with "complex thought process". Complex thought processes

[27] When we introduce a generative theory in chapter 5, we will see this prediction fulfilled. In our discussion of the *qua* problem an explicit answer to how we can ascribe different "aspectual shapes" to wholly nonconscious factors will be given, one without any appeal to potential conscious properties.

have high energy costs and require fairly complex, behaviorally flexible creatures to be of any use. But representation *per se* may prove to be much more like a simple hydrocarbon or primitive amino acid than a well tuned multicellular organism. It may be endemic *to the world,* a building block to be picked up by whatever can use it, not the unique achievement of nervous systems gone jumbo. There may be all sorts of *echoes* in the world, and some of the world's inhabitants can hear them and use them.

This introduces the question of *panpsychism,* the view that from sand grains to human beings, all of the world's objects possess some degree of "mind". "Mind" is usually cashed out as various "higher and lower grades of consciousness" and their attendant thoughts and desires. So we really have two ideas; *panconsciousness* and *pancognition.* Pancognition is almost certainly false. Minds, even very dull minds, are complex activities with relatively high energy costs. So it's fairly clear that most of the world is utterly mindless. Not enough and not the right sorts of things are going on in a sand grain, a carpet stain or a window pane. Panconsciousness is just as problematic if we believe consciousness reflects a highly complex adaptation to world/organism relations. Judging from the behavioral and neurological evidence, conscious beings lose conscious

awareness with fairly subtle interference to their vast neural functions. This may take the form of selective damage--eliminating one's ability to experience an old friend's face as familiar. Or it can be seen in the effects of common drugs or medical anesthesia on global consciousness. Subtle reductions in synaptic and neural performance have drastic, sometimes highly limiting effects on conscious experience. So consciousness, like cognition, appears to be both *complexity* and *dynamism* dependent.

But *panrepresentation*, according to which all the world's objects have nonderivative representational properties, is a notion entirely distinct from pancognition and panconsciousness. Representations needn't be involved in thought processes, nor need they be conscious. Panrepresentation could easily be true--generative theory may christen all sorts of noncognitive entities with representational power. This isn't the absurdity of panpsychism; silicon grit just can't make good use of or in anyway respond to the representational position it may have. Not enough and not the right sorts of things are going on.

Fodor's objection to "crazy pansemanticism"

In a critique of the view that content is simply causal

covariance ("information content"), Fodor offers a simple reduction to absurdity. Many things causally covary in the world--clouds and rain, spots on the skin and measles, smoke and fire. Take the smoke-->fire covariance. Does smoke *represent* fire? Not a chance, Fodor tells us. Because covariance ("carries information about") is transitive, "smoke" (the English word) also covaries with fire (via the sight of smoke) and this would yield the result that "smoke" means *fire.* But it doesn't. It means smoke. Fodor concludes, "There is, in short, a lot less meaning around than there is information."[28]

Is this an objection to panrepresentation? Or merely to an information theory of content? It doesn't follow from what Fodor has said that pansemanticism is false--at best a simple version of information theory is assailed. But I doubt Fodor's objection even threatens this. It is easy to see what is missing. Fodor fails to recognize the simple reply that "smoke" means smoke and, via transitive covariance, *also* fire. There are two contents it enjoys. But they are of a different kind and in Fodor's remarks are conflated. Linguistic meaning and mere representational content are different animals.[29] This shouldn't surprise us. Recalling the case of animals, it is fairly obvious animals

[28] In Fodor's *Theory of Content and Other Essays*, pp. 92-93.
[29] Grice would label the information content "natural" and the socially mediated and generated content "nonnatural".

represent but they hardly do this via linguistic content. At the same time there is no doubt language represents. Both are representational, but of a different kind.

There are other, more traditional points of reply that build on the nature of linguistic representation. Here's a reasonable story: Linguistic meaning focuses on some, not all, covariances. Group processes acting on learning produce this. An information theorist could expand on this by claiming that the use of words involve a number of processes that, in effect, block some contents (like **fire**) in the context of speech communication. Consider the sentence, "The fire is out, but there is still a lot of smoke in the building." On Fodor's simple transitive understanding of information content this sentence would express a contradiction: "The fire's out in the building but there is still a lot of fire in the building." How do we disambiguate the contents of "fire" and "smoke" in these cases? Via a very different pattern of *socially mediated* covariance. The starting point is simple: People don't say "look, smoke!" when in the presence of a flaming, ultra-clean Bunsen burner.

What Fodor needs to show is that an information theorist cannot use the social processes of language learning to produce a kind of content, one still derivative of covariance, that is linguistic in nature.

My preferred lesson is simpler and more interesting. We have encountered an idea that in later chapters we will see is powerful. We have glimpsed the possibility that a representational token might have *multiple contents.*

"As if " verses the real thing

In a related point, Searle introduces the distinction between *real* verses *as-if* content. For Searle a great many things that appear to have representational content actually do not. They only behave "as if" they did. Searle's stock example is the *digestive system.* Searle quotes a passage from the journal *Pharmacology* that strikes him as absurd if taken literally,

> Once the food is past the chrico-pharyngus sphincter, its movement is almost entirely involuntary except for its final expulsion of feces during defecation. *The gastrointestinal tract is a highly intelligent organ that senses* not only the presence of food in the lumen but also its chemical composition, quality, viscosity and also adjusts the rate of propulsion and mixing by producing appropriate patterns of contraction. Due to its *highly developed decision making ability* the gut wall comprised of smooth muscle layers, the neuronal structures and pancrine-endocrine cells *is often called the gut-brain.* [30]

Question: Does the digestive system really represent or,

[30] Quoted by Searle in *Rediscovery of the Mind,* p. 81.

in virtue of its behavior only acts as if it does? To Searle it is a given,

> This is clearly a case of *as-if* intentionality....Does anyone really think there is no principled difference between the gut brain and the real brain?[31]

But the truth is not obvious. Judging from digestion's internal powers of sensitivity and the way these sensitive states are transformed into precise chemical responses and mechanical activities, to me it *seems* fairly likely that it does represent. But again, it is hard to see how we can confidently answer this question without a fairly well developed *generative theory of content.* Searle doesn't derive his skepticism from such a theory. Besides the impression he feels is obvious, he only remarks that if the gut brain is representational, then the whole world is *mental.* But this doesn't follow once we distinguish representation from activities that rely on and use representation,*cognition.* The whole world may indeed be representational. But that says nothing of real mentality--thinking, desiring, believing.

To exclude a system from nonderivative representational power we need some necessary, or probably necessary, condition for this power that it fails to possess. We don't know, *a*

[31] Ibid., p. 81.

priori, what does and doesn't, what can and cannot, nonderivatively represent. Any prospective necessary conditions we have for representation must be set by the generative theory, one that extends to all relevant possible worlds. [32]

If we have a credible theory that allows generative content to extend to the digestive system then so be it. As with *panrepresentation,* this need be no embarrassment. There is little reason to expect the familiar features of human cognition, consciousness, subjectivity or what have you vague but intimate mental property to be necessary. We can't stipulate it with any confidence. Our best explanation of how content is generated is the only reasonable guide to what does and doesn't represent.

Conclusions and basic directions

The main ideas of this chapter are these,

1) Representation is a broad syndrome, one centered on the *aboutness relationship.*

[32] Generative theories only set a special class of sufficient conditions, those relevant to the generation of content or what have you. But if other considerations lead us to believe that these are the *only* conditions for the generation of content (say, in our universe) then we have necessary conditions too (in our universe....).

2) Representation is a reality we cannot reasonably deny.

3) When we aim to *explain* representation we aim to give a ontological, generative explanation.

4) Representation is poorly understood, and is best approached with the broadest, most explanatorily productive outlook possible, the ecumenical attitude.

The direction of research this recommends is simple: *We will try to explain the generation of representation by examining the simplest cases of representational power consistent with the elements of the representational syndrome. This is a decidedly "bottom up" approach, as free of the a priori as is possible.*

Chapter II. Basic Options for a Generative Theory

Where shall we draw our generative explanation from?

The first issue we must face is how we will understand the possibilities of producing "representation". What could make content happen? This is the *source problem*. I believe we should stipulate as little as we can. At this point "representation" should be developed from the only real starting point there is--the concept as we find it in our language and lives. Here is a vast and vague beast, outlined by the "marks" of content. With this in mind, there are two leading approaches to the source problem which we should review before we continue.

The source problem: Internal properties vs. external relations as the content generators

One view of content is very simple: Certain physical things--like regions of brain tissue--simply do represent. They do this as an expression of their *causal powers or other intrinsic properties*. This is the view explored by authors like Cummins,

Hannan and Searle.[33] Searle's version of the view is simple,

What it is to represent a cat, or what have you, is to have particular kinds of causal powers that also exist within the brain.[34]

This view would also be almost vacuous if it were not for the, "*...within the brain.*" caveat. *This introduces an issue of extreme interest and importance.* Searle and friends believe that the brain represents *all by itself.* The presence of a world that acts upon and has certain relationships with the brain is not relevant to the actual production of the representational contents. This is commonly called *internalism* about content. Content is produced internally, by *currently unknown neurological powers.* These are the content generators.

Notice that the internalist isn't claiming a human brain need never interact with its environment for it to enjoy contents that represent the environment. At least in the human case interaction is needed. But this a contingent fact about how

[33] Cummins explores what he calls the "CTC" or *computational theory of mind.* Here the units of content are "data structures", and are wholly intrinsic to the system. See Cummins, *Meaning and Mental Representation* (Cambridge: MIT Press, 1989) chapter 8 and *Representations, Targets and Attitudes* (Cambridge: MIT Press, 1996), chapter 7. Searle and Hannan are squarely within the same tradition. See Hannan's *Subjectivity and Reduction: An introduction to the Mind-Body Problem* (Boulder: Westview Press, 1994) chapter 7 and Searle's *Rediscovery of the Mind* (Cambridge: MIT Press, 1992) and *Intentionality* (Cambridge: Cambridge University Press, 1983).

[34] Essentially this is Searle's theory of the mind, echoed throughout his writing. See, "Minds, Brains and Programs." *Behavior and Brain Sciences 3* (1980), pp. 417-458, for a classic exhibition of it against critics in the A.I. community.

humans develop their internal brain structure and processes. The point of internalism is that these structures and processes create content, not anything about how they came to be. Two identical brains, one that came to have ideas of plants, cats, and so on in the usual way, and one that via some *nearly* impossible co-incidence, appears full blown, possessed of all the same inner structure and processes, would both have the same contents (ignoring certain ego-centric ones). [35]

The source problem for internalism is how a cognitive system has different contents in virtue of the internal properties of its structures and processes. The general form of the approach is,

*Representational process or structure R has the content **C** when R has internal property A.*

What counts as A? Particular *as yet unknown* neurological powers. Though this has a wonderful simplicity to it, the source problem remains unsolved until we can say something about what those powers actually are.

Contrast this to a view that supposes world-to-brain causation, and other relations, are the source of content. It is

[35] The indexical "I" is not molecular-duplicate transitive.

brains (or other cognitive systems) *plus* their relations to the world that generate content. Without these relations, there would be no representing the external world. This is *externalism* about content.

An intuitive example of this idea is found in the difference between real and pseudo (imagined) memory. Suppose you've been to Istanbul and I haven't. Also suppose you and I both have *exactly* the same inner experience when asked if we've ever been to Istanbul. We both appear to recall minarets, Saint Sophia, the smell of unleavened bread baking in the morning and so on. But I've never been there. My memories are pseudo-memories of Istanbul created by some pleasant brain disorder while your memories are real--your memories are *caused by being in Istanbul* during your visit.

This is just the situation the externalist sees many of our representations to be in. It is ultimately because of their connections to the world that they represent what they do. When you think **cat** your cognitive process has the content it does because it was developed through *experiencing* real cats. Cats ultimately stand in a causal relation to it. Complex inner structures and processes are necessary to capture and exploit these relationships, but the relationships ultimately create the content. These are the content generators.

The source problem for externalism is how, in virtue of their relations to the external world, different structures and processes of a cognitive system gain their contents. The general form of the solution is simple: It is in virtue of the external world objects and properties that have various relations to the cognitive system that the system has the contents it does. That is,

*When Representation R has relation A to some object or property B, R has the representational content **B**.*

How we understand A is a key issue for externalism. What sort of relations can count as cases of A? Obviously causal/developmental ones. We'll see some more examples soon and explore this issue in fuller detail at the end of the chapter.

The implications of this can be surprising. Imagine a "molecular duplicate" of yourself. A molecular duplicate is a perfect physical replicate, down to the placement of every atom in your body. As usually understood, externalism implies that yourself and a suddenly produced molecular duplicate of you would share few of the same contents, because the duplicate

would lack the connections to the world you have.[36]

Alternatively, such a duplicate in an environment that was indistinguishable, but actually differed in the nature of one of its particular aspects, would produce a different content because of this external difference. Putnam's famous *twin earth* case illustrates this idea. Putnam is an externalist, claiming the content of a perceptual representation is given by the *causal source.*[37]

Twin earth is a molecular duplicate of Earth, down to perfect molecular duplicates of the people now living on our planet, except for one thing. On *twin earth* the liquid stuff that your molecular duplicate causally interacts with (drinks, swims in, is taught about....) and calls "water" has a different chemical composition than it does here. It's not H_2O, it's "XYZ". Putnam infers that your *twin-earth twin* would have a different representational content when he or she sees "water" because the thing in the world ultimately responsible for that content is of a different kind. On externalism, basic contents replicate the

[36] Later we will see that this generalization can fail in surprising ways. For now merely let it be a way to make vivid the critical role of the external world in generating contents on externalist accounts. It also ushers in the "Swampman" objection to externalism, where a molecular duplicate of you is said to have all or almost all the contents you do, but with virtually none of the external world relations. We will see that this objection is misguided in several ways, but it is a pivotal one in recent debates between externalists and internalists.

[37] See Putnam's classic paper, "The Meaning of Meaning" in his *Mind, Language and Reality: Philosophical Papers, Vol. 2*, pp. 215-271. New York: Cambridge University Press, 1975. A more involved exploration can be found in his *Representation and Reality* (Cambridge: MIT Press, 1988).

nature of their sources. It has a different source for its content so it differs in content.

A. Why Externalism

Externalism as a better foundation for a theory of content

As long as our project is to give a *generative explanation* for representational content we will appeal to external relations whenever they can do the work. We will use an externalist approach. There are three reasons for this,

(1) *The poverty of our present position*

Where a generative explanation of content is concerned, we are wandering in a third world country; we are *poor nomads*. We are very early in this venture, moving into a new and unknown place. The distance to be covered is vast, the task promises to be challenging. At this early stage of knowledge it would be both irresponsible and arrogant to exclude, by any would-be *a priori* stipulation, or any other "conceptual" maneuver, the obvious explanatory opportunity external relations appear to offer (or internal properties might offer, for that matter). Again, this reflects our ecumenical attitude. We have established nothing about the deep nature of

representation. We know only the syndrome. We could not possibly be in a position to know that externalism is false, whatever our intuitions or predilections lead us to believe, or our "thought experiments" appear to indicate.

(2) *The explanatory opportunities of internalism and externalism*

An important class of arguments is critical here: *Explanatory opportunity arguments.* These arguments are critical in the early phases of any new explanatory endeavor. They recommend the most promising direction for precious intellectual effort and research. They also identify probable *dead ends*, approaches with little hope of success. There are three types of opportunity arguments; positive opportunity, negative opportunity and contrast arguments. Positive arguments gesture toward a compelling opportunity for explanation, negative arguments detail why a possible approach is unlikely to succeed, and contrast arguments contain both, pointing to a dramatic contrast in explanatory opportunities between two or more approaches.

These arguments are not *explanatory power* arguments. They don't point to a body of successful explanation that far exceeds a competing approach's and argue that this is ground

for accepting the more powerful approach. They only recommend on the basis of what is already known the best approach for expanding our explanatory power. Much of what passes as explanatory power arguments are really opportunity arguments. The *explanatory power* argument for naturalism over substance dualism has a large opportunity component to it, since naturalist theories of mind are still in their infancy. There is the vivid opportunity suggested by the vast and remarkably subtle correlations between brain processes and mentality. There is virtually no explanatory opportunity to be found in an unknown and apparently unobservable mental substance. So the contrast in opportunity is plain and compelling. There is *no clear explanatory opportunity* to be found in substance dualism. The apparent opportunity in naturalism is opulent by comparison.

The negative explanatory opportunity argument against internalism

Just as with substance dualism, explanatory opportunity arguments weigh heavily against internalism. Again, the source problem for internalism is how, in virtue of their internal properties, different brain processes and structures have the content they do. Here I want to focus on contents that represent

the external world. Where the representation of the external world is concerned, there is virtually *no* explanatory opportunity to be found in internalism. Nervous systems divide into natural scales; the cellular, assemblies of cells, systems of assemblies and the subcortical regions these make up, as well as the cortexes and whole brain.[38] None of these scales, nor any of their combinations, offer the slightest inkling of a way to explain content's generation. A quick review of the levels:

Level 1

The cellular and assembly scales involve molecular exchanges via synaptic projections and formations. The net result of these is the depolarization of adjacent neurons or their failure to depolarize . These are biochemical systems and our understanding of them as such is extensive. What we don't yet fully understand is how their connecting axonal and dendritic structure is controlled through various chemical gradients in the intercellular surround, and how these rely on genetics and previous history of depolarizations and synaptic activation. But these are also biochemical processes operating via biochemical properties, whatever the minutiae that orchestrate them are. There is no reason to suppose that any of this is able

[38] G. M. Shepard. *Neurobiology*. New York: Oxford University Press. 1988. Chapter 1.

to generate a representation of Chairman Mao, my cat or the proposition that the Earth is round. There is no promise in any thing we might expect to learn about these structures and their processes that would include the power to represent the planet Venus or where the car is parked.

Level 2

The system of assemblies level has all these biochemical properties via it's composition. It also has the complexity to produce behavior that can be understood as *computational.* A system of assemblies might be viewed as computing the exclusive "p or q" function, the 2n function or more complex ones.[39] These are the sort of interpretations we also place on artificial computational devices. But the variables within these computations (p, q, m, n and so on) have absolutely no interpretation given by the patterns of neural depolarizations. Understanding the full range of these computational capacities is the stated aim of the neuro-computational branch of cognitive science. The interpretations applied to these variables by cognitive scientists are entirely derived from *external* relations (we will return to this point in the next section). Even granting

[39] A three layer, five unit neural network can compute the exclusive disjunction, for instance. Such a network is described in Judith Dayhoff's *Neural Network Architectures* (New York: Van Nostrand Reinhold, 1990), pp. 76-77.

that systems of assemblies are *intrinsically computational* (which is itself unclear), there is nothing to appeal to within these processes to determine a representational content for them or any variables they may compute. Nor is there any current avenue of research that promises to change this.

Level 3

The sub-cortical, cortical and whole brain scales contain all the properties of the scales they are composed of. They combine to produce the intelligent, profound and just as often quirky and capricious entity we call *the human mind*. A mind without representational powers to extend its view into the world is a horribly impoverished thing--life in a dark closet is a far better fate. But how is this representational reach into the world achieved?

Again, the explanatory opportunity for internalism is virtually nonexistent. All physiological evidence indicates the sub-cortical scale is an elaboration of the systems scale. Systems interact through extensive, apparently statistical patterns of neural stimulation and inhibition. Sub-cortical regions extend this only by vastness, as do entire cortexes. Here we meet the whole brain: Huge waves of depolarization trigger huger waves that divide and disappear in a swirling,

twisting three dimensional *happening* that looks like a scarf tumbling over and over in the wind.

Level 4

Beautiful as it is to behold in high speed, high resolution 3-D cortical images, this fractal swirl is a biochemical, spatial and perhaps vast computational entity. In total isolation from the world, is it anything else? If we believe that our phenomenal experiences (our *qualia;* pure sensations, raw feelings, and so forth without any external world representational component) are caused by this activity, then it is also all of these. But there is no known avenue to attempt the move from patterns of depolarization and synaptic exchanges to a content like **Chairman Mao**. In terms of intrinsic properties we have not moved beyond the system of assemblies scale. These are now possessed by far larger regions of neural tissue, perhaps executing far more complex computational and statistical functions. But there is nothing in *mere vastness* to suggest an intrinsic power to represent the world beyond the brain. So another means of content generation besides "more" must be postulated, one that lurks somewhere in this vastness. And here internalism's explanatory opportunity hits a wall. Current research portrays the cortical and whole brain activities as

either uninterpreted neural processes or via functional interpretations that derive from relations to the external world. Nothing in cognitive science, biopsychology or neurophysiology suggests either the existence or need for a path of departure from this.

The positive opportunity argument for externalism

When we turn our to explanatory opportunity in externalism there are two related themes at work.

The first theme

First is *continuity.* Externalist explanations for mental content have a history of usefulness and success in the cognitive and biological sciences, and there is no reason to suspect a sudden discontinuity here. Externalism's direct solution to the source problem has proven popular. We appear to be on the right track already; future explanations will probably be continuous with our current successes. Research proceeds by noting external situations and observing neurological activities that react or are otherwise involved and related to these. A tiny sampling of externalism at work in the sciences follows.

Brain mapping

In the last twenty years a number of methods of mapping cortical activity have been developed. These include high resolution methods like the P.E.T. (positron emission tomography) scan. In research using P.E.T. scans the method and inferences are transparent in their externalism. Consider shape recognition and manipulation tests; shapes are presented a subject and while she works through them her brain is continually scanned, and areas of activity displayed on a monitor. Assuming that cognition is an *activity*, the researchers conclude they have discovered areas in the subject that understand and solve certain intellectual tasks involving the recognition and representation of shape. Why are these regions said to be shape representing and task-solving regions? Because shapes are what was presented to the subjects' sensory system--real shapes out in the world. There is no intervening theory about the content generating properties intrinsic to these regions. When the processes within these regions are studied the pattern of external presentation and mapping only repeats itself, on ever smaller scales until we reach micro-electrode, single cell recordings. These shapes serve as externalist content generators. There are thousands of similar examples in brain mapping and cortical function

research.

Behavioral ecology

This branch of biology studies evolutionary explanations for behavioral dispositions. The Bee Eater is a bird that nests in vast, cooperative colonies and is famous for its helping behavior. Younger, nonreproductive birds will help feed the offspring of mated pairs. Why has this behavior evolved, since it doesn't appear to aid in the survival or reproductive success of the helpers? A leading hypothesis involves the birds representing the degree of genetic relatedness of other birds in the colony, and helping in accordance with degree of relatedness. This benefits reproductive success in terms of the amount of genetic material shared by the helpers and the offspring of other mated pairs that makes it into the next generation. How do the birds "know" who is closer kin? A number of external clues exist in terms of scent, markings and subtle behaviors. These *indicate* relatedness. Indication is a relation that holds between external entities and certain internal cognitive states of the birds. It is an externalist generator. Further, the helpers must represent which birds are associated with which clues, especially when these are behavioral clues. Naturally, the helpers learn to represent mated pairs by

causally interacting with them--another apparent externalist generator. There are thousands of similar examples in behavioral ecology.

Clinical psychology

Among the thousands of disorders treated in clinical psychology exists a biological/behavioral pattern called *attachment disorder.* Attachment disorder occurs when infants are deprived of care giving tactile and auditory stimulation. These crib-children have a high probability of later developing little sense of compassion or emotional connection with other individuals and feeling beings. They torture pets, murder siblings, commit acts of arson calculated to maim and kill, and so on. They are as lethal in their acts as they are bizarre in their lack of regret.

The current explanation is based on the model of mammalian mother/infant bonding. Because these children failed to "bond" with a parental figure during an innately sensitive period, they lacked the basic representations of compassion and care giving that most humans receive and are innately predisposed to then develop. What makes the experiences of a newborn that does undergo bonding representations of compassion and care giving? These are the

types of behaviors of his mother exposes him to, and he is causally effected by them. This is an externalist generator. There are thousands of similar examples in clinical psychology.

Second theme

The second theme is *natural fit* with biological evolution. Externalism is exactly what we should expect to be at work in world-bound cognitive systems. If we wanted to build systems with representational content, an obvious avenue would be to tap the rich bank account of the world. That is, why not just build the representation of the furniture of the world in large part with the furniture of the world itself? *Let the world import its individuals and kinds to the brain.* Content is latent in what already exists outside the mind, the mind just needs to be able to access and use that latency. The relations it has to the world are the obvious means of access. *This is a much simpler and direct approach than evolving systems that first produce external-world contents, de novo it seems, and then find things in the world to use these on.*

This natural fit extends to artificial cognitive systems (A.I. systems). With the advent of high speed linear processors like those found in high-end desk top computers and workstations, and the development of powerful neural network emulations for

these processors, we now appear to have the ability to produce realtime learners and problem solving systems. Especially where neural networks are concerned, we have the challenge of how to understand them representationally. Externalism is the obvious solution. Derivative solutions ("what does the programmer *intend* the A.I.'s system's states to represent") are especially inadequate in the case of neural networks, because these learn via their own inner dynamic and often it is unclear just what exactly they are learning until we carefully examine the relations between their environmental input, hidden layers and outputs.These systems represent in terms of their relations to the objects and types that their learning processes react to. Indeed, there seems no other way to understand their representational properties.

Externalism's *natural fit* with design will only become more and more important with time. There is a great deal of work to do to. How all this will be accomplished isn't known--but the opportunity is *manifest*.

In both the progress of recent research and from an evolutionary/design stance, the explanatory opportunities in externalism are excellent--*the* promising avenue for the development of a generative theory of representational content.

The contrast argument

This is just the contrast struck by the two previous sections. It's *short and sweet.* There is no apparent avenue for the development of internalism, externalism has many avenues and immediately solves the basic issue in content generation, the source problem. Further, externalism is being pursued and developed with great success in the representational sciences. The theory guiding conclusion is simple: The explanatory opportunity in externalism is vastly superior at this point to internalism's. We should pursue an externalist theory at this time.

(3) *The improbability of an accurate internally constructed representational scheme*

I think it is critical to drive home the lesson of the negative opportunity argument. Let's focus on an *example* of internalism's fundamental puzzlement with the*source problem.*

I want to consider a problem for internalism I call the *matching problem.* The problem is how internalism can account for the fact we develop the correct representations for the variety of ordinary objects and properties in the world. Our representations *match up.* It's important to distinguish this from the problem of how we are able to adequately behave in the

world. This problem is much more tractable for internalism because for the most part I believe we *can* provide explanations for behavior (narrowly construed) without any recourse to representation.[40] The behavioral problem can be easily solved by appeal to various input/output properties of neural networks. The problem for internalism is how representation is correctly deployed.

Recall internalism's basic claim,

The brain generates contents of the external world in virtue of its internal properties. Relations to the external world are not necessarily required to do this.

Things go wrong when we ask how it is that the contents the brain brews up are so remarkably, wonderfully appropriate to the world. The fit is incredible.[41] It seems the brain *must* rely on content produced by external world relations. This isn't a point about the Earthly need for environmental stimulus to build a cognitive system, it's the problem of matching appropriate

[40] This is not to say content based explanations for behavior are not also valuable, and often far more so. Indeed, for externalism the problem of accurate representations more or less is solved along with the behavioral problem, as should become plain by the end of this work.

[41] *At least so it seems*....In certain contexts the possibility of systematic *misfit* are notoriously interesting, for instance the selective advantages of suffering severe *paranoia* in many prey-species. Soldiers in combat quickly develop a "hit the dirt" reflex to the sound of explosions, only to be plagued after years of peace time by the same behavior in the presence of fireworks, car back-fires and other loud noises.

contents to this stimulus, in the absence of this content being *supplied* by the stimulus.

The initial problem for an internalist content-generator which has to represent an external world is that there is an infinite set of possible objects, processes, properties and other features, but it needs to generate contents that represent a tiny subset of these; the *actual ones*. Our world contains the smallest fraction of this infinite set (recall how many infinite is).Without relying on the world to produce the appropriate representations--representations of *its* entities--how do we explain the almost impossible good fortune of having developed representations of its furniture and very little that isn't in some way reflected in the world --the way a unicorn is reflected in real horns and horses? Unlike externalism, which ultimately relies on the world providing the content, internalism faces an initial and difficult *matching problem*.

Because most of these representational contents are in some way rooted in experience the brain must relate them to perceptual input. Consider visual input. Here we are presented with patches of color, their levels of illumination and boundaries. These have to be mapped onto internally produced representations of the objects that stand behind them. But a dilemma immediately appears: Either (1) the brain does this by

recognizing which patches are what kind of thing; *the input is matched to the correct preexisting content,* or (2) it receives the input and then *constructs a representation of the thing it is caused by.* Neither of these can succeed. The former is a nonstarter because it entails that we never learn representations. But it is obvious that either via evolution or more typically, experience, we do. Both the former and the latter are improbable because the brain would have to know what kind of thing (external world feature, property or object) produced the input in order to construct or assign a representation that maps on to the *right* object. Remember, according to internalism the content of being the object in question is not carried in the input. Either the brain must guess what sort of object or feature in the world is the causal agent, or it has some prior knowledge of what the right content should be. If it guesses the odds of success are some finite number over infinity--the poorest imaginable. If it already knows the circularity is plain: To know what the thing is we already have to have a representation of it--which either launches a vicious regress of learning episodes like the one just described or returns us to the former option (1) and the denial of content learning.

Again, this isn't a question behavior can answer--even

uttering in the right circumstances the sounds, "I-see-a-cow." Neural networks can quickly and easily acquire the capacity to make such noises in a given circumstance, say when a cow appears. They needn't have the content **cow** to do this.[42]

But imagine a network acquires the content **cow**. (Given the right scenario, I'm sure they can.) For an externalist this is naturally explained with the kinds of entities that produce the external world inputs that the network discriminates--like inputs ultimately traceable back to real cows. If the system lacks the needed relations, its behavioral declarations are not reflective of contents. But how does internalism determine the right answer here? For all the internalist can show, every time the system emits the sound we write as "cow", it has activated a representation with the content **Richard Nixon**. Though a network may systematically respond a certain way to the input of cows there is nothing in this that allows the internalist to therefore claim the content is **cow**. Far from it, for this is the externalist explanation; relations (perhaps indicative or causal) to the actual cows are the root source of the content of **cow**.

True, the externalist has more she needs to say. In coming chapters I hope to show just what. But the key difficulty--getting an appropriate content into the cognitive system--has

[42] At least they don't do this *in virtue* of having this or any other conceptual content.

been tackled. Internalism's work is stalled, with no apparent way to get started--which doesn't leave it far to go on the way to failure.

Internalism and learning

A possible way out of the above pinch: *Maybe the objection here is only that the internal content generator must start off badly. Perhaps it can catch up by progressively changing its representations, ultimately approaching an accurate mapping. Perhaps it can learn the right contents.* This automatically sets aside (1). Given the extremity of the situation in (2), this is highly unlikely to ever succeed--but an exploration of it helps to simultaneously drive home how dire the matching problem is for internalism and how much better a position externalism is in.

The matching problem is only compounded when the brain attempts to make modifications to representational contents. The brain changes the structure of our representations. It fine tunes them, elaborating and correcting. Two problems arise: (1) Why it would, since it behavior is congruent with the environment, is a mystery. If our behavioral responses are well tuned to the world, but radically incorrect contents are generated for them, why should there be any

occasion for the activity of correction? Everything works. So even if our conceptual life is schizophrenic, why bother changing it?

But let's suppose for whatever reason the brain makes the attempt. Enter our second problem: (2) *How* does the brain make the attempt at change? Consider a basic kind of correction: Imagine the brain maps the concept **wheel** onto spheres. There are important differences of course; wheels have only a single symmetric axis of rotation passing at a right angle, through the center of the plane. Spheres have as many exactly similar axis as you like, all passing through the center point. The brain must discover at least this difference. But there are any number of possible shapes in a continual transformation from wheel to sphere. Recall how difficult these discriminations are. Just as John Locke predicted centuries ago, blind individuals when suddenly given their sight cannot visually distinguish circles from triangles, much less from spheres. The differences are *not given by* perceptual input alone, even though for externalism they are *represented in* the input. How does the brain discover which ways the content **wheel** should be modified to become **sphere**?

Some would-be solutions allow us to detect the pattern of failure,

1) A trial and error approach, where certain modifications are made *randomly*, seems the only answer not tainted with something like *previous knowledge* of the correct solution. Guessing is hopeless, which means a true trial and error approach is going, in all odds, to continue unsuccessfully until death. A more advanced idea: Perhaps the random approach could be linked with a preprocessing of sorts. Contents would be previewed in terms of their implications for visual input and then their implications for patterns of visual input considered. But this is just another form of guessing. It also reinstates the problem when we ask how the brain knows what, "the implications for the patterns of visual input...." might be. How does the internal content generator know which contents have the sort of implications for input that the real world objects and properties do? We are back to our original guessing game, with virtually no progress made against the infinite set of possibilities that must be explored.

2) Perhaps the brain can exploit the successful behavioral repertoire to repair its mistaken contents. But these neural structures have no apparent external-world content, unless we are viewing them via externalism. They are merely input/output transformation devices. They operate the same regardless of the environment, as long as the perception input is the same. If

we put them in a virtual reality setting, nothing changes. Imagine a circle/sphere discriminator; from the view of internalism the discriminations are uninterpreted; x and y. As it happens, x corresponds nicely with external world objects of the kind circle, and y, spheres. But this is a fact of external world relations, not anything within the network. The network has nothing to offer the internal content generator. Nor is there any evolutionary reason to suppose it would offer any guidance if it could. No selectable behavior would benefit from this.

3) There is a sense in which the input/output transformations of neural networks *resemble* the things in the environment they are reacting to. "Resemblance" is a fairly open ended notion. Perhaps it can accommodate this: There is a series of translations that lead us from the output, back through the network, to the inputs, from the input to the causal vector (light, sonic vibrations, etc.) to the causal source object. The x and y discriminations can be said via the application of this translation, to abstractly resemble the things in the world they discriminate. Suppose we let the brain's internal content generator *represent* these x/y discriminations. Then it represents something that in a very abstract sense resembles entities in the world. Let this representation be another step in the translation. Then the representation itself has a

resemblance relation to the things in the world. This could be used, in the fashion of 18th century empiricists, to launch an explanation of external world content: These internally generated contents are about external world entities, because they *resemble* these in a highly abstract, translation driven sense. So as the network improves its discrimination abilities, so the internalist generator improves its resemblance, and hence accuracy.

The problem here is the same in (2) of course. This is an interesting account, one an externalist might be interested in using parts of. But internalism is barred from it: The resemblance relation is ultimately a relation to the external world, and it is this external-world relation that drives the entire derivation of content. Remove the external world and the scheme collapses. This is an externalist account.

The diagnosis in all these cases is the same. The internal content generator is isolated from the world's latent contents. It's in the dark and can only grope blindly. Which is just our original matching problem. The root of the matching problem for internalism? It's fairly plain. Internalism has no inkling, and no way of getting an inkling, of what a working solution to the *source problem* might be.

B. Psychological Ecology

External bases for content; representation and psychological ecology

We are going to develop an externalist, generative theory of content. We are working on a task only recently confronted by contemporary philosophy. I believe that only a misguided pretense to knowledge, or a related and just as misguided *a priorism*, would have us set aside any apparent bases for the generation of content. I suggest a opportunistic approach. Within the constraints of the physical, our present state of understanding provides virtually no prior constraints on the resources for good theory. Indeed, the *general conception* of representation I mean to explore will leave us with few if any external relations denied the power of generating content in a cognitive system.

I call this general conception *representation as psychological ecology*. Psychological ecology sees representation as an extension of the world into an organism, via the relations and other properties its cognition bear to the larger physical world. Psychological ecology is lush in its Naturalism. Representational systems are ones who's internal states are in various ways, via various processes and relations,

reflective of the processes, properties and objects of the world. To represent is to occupy a certain multidimensional *position* in the world. Our psychology's content is an *ecological* fact, much the way our larger biological nature is. Two particularly powerful ideas inform psychological ecology,

1) *Holism of the world.* Long a metaphysical and even religious notion, the idea that the world is a interconnected *network* or *web of being* is as ancient as it is pregnant. Today this idea has received new impetus from a variety of sources, including Quantum Mechanics, Cosmology and biological Ecology. Representation expresses the membership of mentality in this greater whole. It is the cognitive solvent of our isolation.

2) *Recapitulation, self-similarity and replication.* Endless parallels in nature's geometry, ever-present, overwhelming mathematical analogies between diverse processes and structures, a constant repetition of structure and events in time--all these characterize the world on all scales and within and between all its modes of existence. The world is a vast collection of mutual echoes. Representation easily fits this fundamental aspect. Here *minds* are the stage upon which reality duplicates itself.

Content consists of relationships between cognitive systems and the world, and these relationships constitute a kind of replication or reoccurrence of the world within the cognitive system. This broad, minimalist image is the natural child of the ecumenical attitude: *Many things might be nonderivative representations. This might occur in any number of ways.*

Particulars bare these abstract themes well: A tree on a hillside is a vast cluster of ecological facts: Its lean, the grappling of its root system with the rocks on and beneath the soil, the shape and sky ward tilt of its leaves, the length and array of its branches, the timing of its blooms, the thickness and flow of its sap, the number of rings beneath its bark; all of this intimately represents the highly complex ecology of the tree. All these reflect both the themes of holism and recapitulation. The tree is embedded within all sorts of processes. These express themselves in the tree in a way that reveals their existence by repeating aspects of their form. These are what we sometimes call *natural signs*--the proverbial, "Where's the smoke there's fire.". But they are so much more subtle and revealing than we might think. The perspective of psychological ecology extends natural signs and other relations into the cognitive realm, combines them with their *manipulation and transformation,* and

creates cognitive representation.

Here the holism of the world is expressed through the intimate connection between a cognitive system and the world that representation creates. It is a holism cognitive systems consume enormous resources to expand and maintain. The stuff of this connectedness is found in the vast *repetition* and *transformation* that representation makes of the world's furnishings. We witness an ever greater, ever tighter pattern of *world re-creation* in the most complex cognitive systems. We find a vast and ever growing *psychological* ecology to be found in the relationship between representers and their world. Representations are what the cognitive side of this relationship is composed of. When we ascribe content we are really only making note of these relationships.

The fit with Naturalism couldn't be better.

The relations that build a psychological ecology

If to represent is to occupy a certain "multidimensional position" in the web of the world, what are the dimensions involved? Judging from the complexity of ecologies in general, I think psychological ecology should recommend a *broad* view of the relations that can build its representations. On the issue of external content generation, the ecumenical attitude is this: If a

relation can further extend the holism and repetition that underlies a psychological ecology, it has its place as a content generator. There are a number of *prima facia* world/cognitive-system relations that are representational in a wholly nonderivative sense. Why are they nonderivative content generators? Because each, alone, appears able to reproduce the features of the syndrome of aboutness. There are three major generators I'd like to focus on,[43]

1) Causal connection
2) Indicative covariance
3) Resemblance

(1) and (2) are currently very popular.[44] They are easily confused, since (1) can often ensure (2). But there is no need for x to cause y for y to co-vary with x. The crocus blooms in the Northern hemisphere co-vary very nicely with the onset of low

[43] A fourth, inferential role, is popular in combination with causal or indicative relations. I will not focus on this because it seems to presuppose a solution to the source problem. We ask, "what is the source of the contents that are deployed in various possible inferential roles?". Thus, we are hunting for more basic units. *This* is a problem that the relations above directly address.

[44] Devitt and Sterelny (in Devitt, Michael, and Sterelny, Kim, *Language and Reality: An introduction to the Philosophy of Language*. Cambridge: MIT Press, 1987), Putnam (in *Representation and Reality*, Cambridge: MIT Press, 1988) and Fodor (in *Psychosemantics: The Problem of Meaning in Philosophy of Mind*, Cambridge: MIT Press, 1987) have each pursued this project. An early attempt was Armstrong's ground breaking work on a "Causal Theory of Mind"(reprinted in Lycan's *Matter and Consciousness*).

altitude snowfall in the Southern hemisphere, but they are not caused by it. (3) is the classic source of 18th century empiricism, a centerpiece in the tradition of Berkeley, Locke and Hume.[45]

On an open view of content generators, all three of these can generate content. Recall the *syndrome*, a suggestive pattern of abilities or powers,

The first is operational,

Things with content can be used by a cognitive system to act upon the world in reliable ways--they allow the system to control the world beyond itself in a way that adapts itself to the facts of the world; past, present and most interestingly, future.

The second, related mark is intrinsic,

Content can be (and almost always is) in some sense separate or independent from the thing it represents. First, content has an aspect of autonomy from reality--it can be mistaken or fictitious , involve what no longer or doesn't yet exist, but still be

[45]This view is explicit in Berkeley's claim that an idea can be like nothing but another idea--the foundation of his classic, *A Treatise Concerning the Principles of Understanding* (reprinted in *George Berkeley: Principles, Dialogues, and Philosophical Correspondence*, Colin Murray Turbayne, editor, New York:Macmillan Publishing). Locke and Hume also continually speak of representations resembling what they represent.

contentful, because it involves something beyond itself, and this something may or may not exist in the world. Second, content manages (paradoxically) to be what it is not: My thought about Little Cat B is not itself a cat. But I can consider and image Little Cat B by attending to my thoughts about him. Content is also most always separate from the representation that has it--this dissertation is about content, not about dissertations, the sentence, "Cats are gods." is about cats, not sentences.

The third is experiential,

Content is something we appear to encounter in the way we experience our lives and relations to the world. We experience things about our world. We apparently have perceptions, thoughts and hopes. They are about our world and lives. We experience learning what things are and how things work through our representations of them. We experience using representations to make (at least apparent) contact with reality.

All three of these relations fit well with the first mark of content; causal connection, indication and resemblance can each be used to manage our world representationally. Causal connections and indications alert us to the most essential task

of representation: the presence of the changes in our environment. It is really no different than the representational significance of the phone ringing. Resemblance is obviously useful for control. Photos, maps, and mathematical models all can be used in scenarios where their features resemble a subject matter we aim to use or alter.

The second mark, while a bit more abstract and paradoxical in nature, is also easily seen in systems that engage these three relations to the environment. Causal, indicative and resemblance relations allow for the two fold autonomy from the thing represented and the token doing the representation, because they are neither of these, but act as a *via media* between the two. Our explanation of error will have to wait for chapter 4, but there we will build an account that can be used with all three relations.

The third mark must not be taken too ambitiously. We may not directly encounter the fact that causal connections, indication or resemblance is at work in providing the content of our tokens. But what we do encounter in our awareness is something I believe can be understood by appeal to these relations. This is, after all, part of the task of a generative theory.

So it looks like there is a good *prima facia* case for appealing to these three relations. How far we can take them in

our explanation remains to be seen, but we've got our hands on some promising tools.

Causal relations seem especially important in perception and learning. It's because Europeans of the 16th century had causally interacted with ("seen, felt, tasted") *ice* that they had a representation of ice, but certain royalty on the Indian subcontinent at the same time were incredulous at the Europeans' claims of *water that was solid when very cold.* They had never seen (etc.) the stuff. Causal relations seem to have powerful application to reference. In philosophy of language they have been used very effectively by writers like Kripke.[46] In mental representation they have obvious application to episodes where something is perceived but its features are not distinguished in a way that allows us to type-classify it.

Indication also has a straightforward and powerful appeal. I have given a indicative cipher for the low altitude snowfall in the Southern hemisphere--crocus blooms. You can indicate one event with the other (and vise versa). Crocus blooms can, in concert with some interpretive machinery, play the central role in generating a representation of when the snow falls Down Under. This isn't just an epistemological point about *accurately*

[46] See Kripke's classic, *Naming and Necessity.*

or *reliably* tracking Southern snowfalls. Indication needn't be certain, never actually is, and can be probable to a *very low* degree. It can be highly inaccurate but still representational. If we are intently looking for something--say a $100 bill we've misplaced--lots of paper edges of such width, green bits of this and that shade, patterns of print of a certain general sort and so on will make us startle with expectation, "There it is!". They all are *very* low probability indicators. In a heightened (read: *semi-desparate*) level of awareness these indicators' representational capacities are all recruited to help locate U.S. currency. Indication properties--even if epistemically pitiful--appear to be representational in their own right. This can be extended to the point where the odds of the indication are 1/first order infinity. This epistemic residue is simply the consequence of the fact that for R to indicate A, it must co-occur with A at least once. But once it does, this need never happen again for R to represent A. We can imagine scenarios where R can only co-occur with A once, and never again can the twain meet. To the unwary cognizer, R can still indicate A.

Indication is obviously reference-like, since the fact a representation indicates something needn't in anyway reveal the properties that characterize that object (beyond the fact it is indicated).

Resemblance: A classic content generator returns from exile.

Resemblance is a highly intuitive but recently much maligned generator. Many Naturalists believe the only respectable relations to generate content from are causal and/or indicative.[47] I disagree. Take these two, causally speaking, very different processes,

Process (1): Joe, looking at an image produced by a turtle algorithm designed and carefully crafted by a student of botany and computer science, to produce an image of a lilac.

Process (2): Jenny, witnessing the same image as produced by the same turtle algorithm, but this time randomly generated by her computer. She is without any ability to relate it as a whole to other representations she has--suppose she has never seen, heard of or has any other inkling of *flower.*

Joe sees a Lilac. Does Jenny? On an ecumenical approach the answer can be "Yes". This isn't a matter of causal chain limits, because Jenny's image has no Lilac at the limits of the chain. Instead, Jenny represents a Lilac because process (2) falls under the process type of *being a process that includes*

[47] Devitt is a vocal example.

the existence of a pattern that has the same spatial properties as a pattern that is caused by Lilacs. This is a *resemblance* relation. Because Lilacs appear as a feature of process (2), they can be used to specify a content.

This is contrary to the basic tenet of causal content theorists like Putnam. His famous "Ant and Winston Churchill" story illustrates his view,

> Imagine a busy, somewhat confused ant crawling in the sand. As she crawls about her vast path, she traces a complex pattern of lines onto the sand. We see these and immediately realize that they look exactly like an excellent drawing of Winston Churchill. But is the ant trail *itself* a picture *of* Winston? Has the ant produced a representation of Winston? Or do we merely choose to *interpret* it that way? Does the trail by itself represent Winston?[48]

For Putnam the answer is a simple "No". Because the ant's crawling about in that way had no *causal link* to Churchill--Churchill is not the limit of any causal chain the drawing is part of, there is no representation of Winston. Instead it was entirely fortuitous that the trace happened to look like Winston to us. We merely interpret it as being *of* Winston. But if it had been caused

[48] Putnam discusses this kind of case in his "The Meaning of 'Meaning'", in Putnam, Hilary, *Mind, Language, and Reality: Philosophical Papers*, Vol. 2. New York: Cambridge University Press, 1975.

by the actual likeness of Winston (and perhaps *caused in the right way*, the way a recording of his voice, a photo of his face or a memory of his courage is) then it would be *about* Winston, and not merely something we can make-believe is.

Whatever our intuitions about such a case, they are not sacrosanct. There is no reason why we should have special, authoritive insight into these issues.[49] Besides, intuitions differ in this case (myself and many others don't have Putnam's intuition) and we can't *all* be right. So process relativity wants to know the *types* of physical processes involved in the Lilac case, or in the Ant story to see if Putnam's intuitions are borne out. It looks like they are not. Recall that the same instance of a process can fall under different types. In the Lilac case there is a common physical type, having the same spatial relations of parts that the parts of the other has.[50] And in both cases, these are the spatial relations that a Lilac has.

Understanding "resemblance"

There are two leading criticisms of resemblance as a generator. First, it's plain that our "ideas" do not resemble what they represent. Our idea of intelligence needn't be all that

[49] Neither in evolution, our cultural or personal experience is there any reason to believe we have developed the ability to have accurate *intuitions* about matters as obscure as the ultimate bases of content generation.

[50] This physical type may be the sort of fact represented by a statistical gradient of Lilac-like features.

bright, our idea of red is not itself a red thing, our idea of a black hole is not itself of zero volume and infinite density, or of any volume or density, and so on. The second typical worry focuses on the view that ideas are *pictorial*--that they picture x, and so resemble x, and so represent x. The criticism is that given a mental picture of The Chairman Mao, we can't say what exactly it represents. We have (if done correctly) pictured the great leader himself, his very proletarian cotton quilted suit, his protective, piercing eyes, his stately posture and his self-assured, revolutionary countenance. *Which* of these is represented? Which of these do we select as *the* content?

This second concern is an obvious incarnation of what I call the *choice problem*. The criticism is set aside by the rejection of the *choice problem*. It is premature to discuss the problem at length now but it will become a leading theme in later chapters. For now I will simply say that the basic solution is straightforward. A picture of The Chairman represents *all* these virtuous things and more. But these different contents can be *segregated* from one another because of fundamental differences in their generating processes and/or relations. We will return to this topic at length in chapters 3 and 4. Until then, let's set this aside and return to the first worry: Our representations don't resemble what they are about.

Ways of resemblance

The criticism that representations don't resemble what they are about is well taken but it presupposes a far more narrow view of resemblance than we might. Resemblance is a very open ended relation. In its *basic* form, resemblance exists whenever we have the following,

There is some property, P, that both X and Y share.

X and Y resemble in terms of P. P is the bridge of their resemblance. One P in a field of many can be a challenge to recognize. IQ tests exploit this. Pile up enough subtle ones and the effect can be as overwhelming as it is hard to satisfactorily express: The resemblance, not merely *reminding,* I find between some of Mozart's music and certain alpine meadows comes to my mind.

I believe there is no such thing as resemblance unless it is in terms of some shared properties. Properties can resemble on this scheme too, if they are complex ones (properties made up of other (often simpler) ones). So it is that the property of being circular and being square resemble in terms of having *points on their perimeters at an equal distance from the center*--a

complex property involving others. An octagon more closely resembles a circle in this respect, because it has more of these points than a square does, so more closely approaches the infinite number a circle has.

The question is how obscure can *P* be and still constitute a bridge of resemblance? Here's a resemblance bridge that is a a bit trickier still: The property of having the same proportion hold between temporal beats and spatial distances on a line. Here the resemblance involves a *translation property*. It translates (in the nonsemantic sense of "converts") a spatial to a temporal proportion, and vise versa. Perhaps what is meant by "resemblance" is really "recovery". It is a simple resemblance bridge via the shared proportion that defines the translation.This property seems as physical as one could be, since the proportion has physical, causal effects, like when the distances are bumps on a roll of paper used in old fashion self-playing pianos, or the magnetic tape in your tape deck. A far more complex translation resemblance exists between a material object's mass, velocity, shape, composition and the impact shape it makes on a surface. This fairly obvious in an old style typewriter, less so but just as real in a meteor crater.

Unlike shape or color resemblances, translation resemblances are just the sort of thing that can be far from

manifest. There seems no in-principle limit to their complexity, (as long as they are real properties, which my ontological preferences identify with the physical), and so no telling how obscure they might be to human observers. Indeed, the basic resemblance of being of equal volume (tall glass verses short) easily eludes even adults. Translation resemblances are just the sort of thing that avoids the general criticism above. *Perhaps* our representation of a black hole does have a highly sophisticated translation resemblance to the real thing.

There seems to be at least two kinds of translation resemblance. The first is fully reversible translation; the steps that take us from input to output form can be repeated in reverse order to return us to the original input.The second is irreversible translations.

Computer graphics programs like Adobe Photoshop include an image encryption process. It takes a spatial distribution of colored pixels and produces a string of code. This is a *reversible* translation. An image is reduced to Postscript code--a bizarre stretch of text-like symbols--and the same operations in reverse order can reproduce an exactly similar image. A classification network is an example of *irreversible* translation. Imagine a network where tree images are classified into *species*. The network starts with an image on a plane as

input, uses a hidden layer to transform the input to an output that specifies the tree type. A representational explanation starts with causal relations: Tree images are caused by trees. Then we use resemblance to address the output: These images are mapped onto species-classes and degree of fit is revealed. Why are these classes representative of some fact or feature of trees, in particular *tree species?* One perfectly acceptable answer is that there is a *translation resemblance* from the species revealing features of real trees to their species descriptions. The hidden layer performs this translation. But from the output, "Douglas Fir" we cannot recover the original image in the kind of detail that would distinguish pictures of two different looking Douglas Firs. We would get, if anything, a sort of prototypical looking Fir.

Where do translation resemblances appear in human representation? They are everywhere in the sciences, most obviously when we encounter *models*. Take fractal dimension models for certain types of geological features--continents and mountains, for instance. A computational system that produces a stream of values according to a given fractal dimension can have a translation resemblance to the pattern of peaks, cliffs, gorges and valleys of the Alps. Contemporary, nonlinear models of population change also have such a translation

between a fractal dimension and herds of caribou.[51]

Translation resemblances are also *one* way to understand the representational capacities of neural networks, particularly their enigmatic *hidden layers*.[52] There is a translation resemblance between the inputs and outputs of a network; this resemblance property is at work in the the dynamic profile of the hidden layer. Because the inputs and output have this translation resemblance, the latter can be said to represent by translation resemblance the former, or to some *more definite, selective fact* about the former. Which of these scenarios, direct representation or selective representation, depends on the nature of the translation.

Translation resemblances are important candidates for content generators in cognitive systems. They are clearly Naturalistic in kind. For instance, they can be understood as a *process*, whose steps are wholly expressed in terms of Natural variables, where by the representational content is converted into some pattern of features or other properties the represented entity also has. They appear to be able to account for description, since the translation can be attuned to many of

[51] An excellent review of these resemblances can be found in Manfred Schroder's *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise* (New York: W. H. Freeman and Company, 1991).

[52] Paul Churchland provides a highly detailed account of resemblance properties in diverse neural networks in his "Conceptual Similarity Across Sensory and Neural Diversity: The Fodor/Lepore Challenge Answered.", *Journal of Philosophy*, Vol. XCV (1998): pp. 5-32.

the properties of the entity represented. (Description might also be composited from simpler representations that do not derive their content from resemblance, but resemblance provides a more direct descriptive representation.)[53]

Externalism revisited; the swampman finds a home with externalists

I would like to comment on one interesting consequence of taking translation resemblances seriously. It provides a surprising reply to the "Swampman" objection to externalist solutions to the source problem. The *Swampman* is a swamp goo, gas and muck miracle--out of the ick monsters are born, and some of them might be molecular duplicates of *you*. Or so goes the story. Now such a duplicate has enjoyed virtually none of the causal contact history, subsequent learning history or species evolution that inform your being. That is, *Swampman has extremely limited external world relations compared to yours. He has no learning history, no evolutionary past, nothing of the causal connections that characterize your representations.* But Swampman, being a duplicate of you, enjoys all your inner mental states. He (or she) *feels* just like you do. Seems to remember every little thing you do. Apparently

[53] It is interesting that in many cases of description, "translation resemblance" gives us a new way of articulating the correspondence theory of truth.

knows everything you do. So doesn't Swampman enjoy *all the mental contents* you do? And doesn't he know this as surely as you do? But this is accomplished without so many of the external world relations you enjoy. So these must not play the generative role we might believe they do; such contents are generated by the internal features of the brain alone--the thesis of *internalism. In contemporary terms, Swampman shows that our contents supervene on internal properties only.*

The swampman has widely been taken to be a supreme embarrassment for externalist accounts. It needn't be. Swampman internally experiences all we do, in terms of the inner character of his or her conscious experience. But are these conscious states representative of all the things they are in us? We can't say, just by looking at what it *consciously feels,* etc., to be Swampman. The objection rests on a false assumption--that we have direct access to the contents of our mental states.There is no reason why we should, nor is it clear on *either* externalism *or* internalism how we could. For internalism's part, how exactly are we to know what internal constellations of properties represent. Their mere presence and our awareness of them does nothing to answer this question. But let's set this deeper point aside. The claim is that Swampman would have the same representational states,

contra externalism.

The multiple content-generative bases for content Naturalism and the psychological ecology model it encourages offer us another reply. These invest swampman with a great deal of contents, including many of the very same ones *we* have. This is especially obvious when we consider resemblance translation contents. Swampman's neural memory array has the same translation resemblances to perceptual objects that yours does. This includes all of your perceptual representations, as well as a great deal more.

I introduce this point only to show how multiple bases for contents and their implications can complicate apparently straightforward issues between internalism and externalism. Once we admit multiple bases, our explanatory resources expand in ways we have only the slightest inkling of.[54]

C. The Choice Problem

Multiple bases for content and the choice problem

The main apparent problem for a view that allows so many possible content generators is that it gets *too much*

[54] A similar point may help defuse the epistemic objection to externalism: We know our contents of our representations, but we don't know the external relations/extended processes that are claimed to generate them. Knowledge of a translation resemblance may be very much *like* knowledge of the object in the world that is represented.

accomplished. The same representation has multiple contents. Thus, *the choice problem.* Here's a typical expression of the choice problem *as* a problem,

> ....we shall call [it] the "conjunction problem"....The standard example used to illustrate the problem is that in the frog's natural environment, when it zaps a bug it also zaps a black dot. What does S [the frog's representation of the zapped bug] mean? [Bug or black dot?][55]

The authors see this as a serious problem and attempt to pin down a single content. There are other, more complex, examples too: A brain sees crocuses and forms a visual cortex representation of them. Causally this is understood to include the content **crocus**, but indicatively it also has the content **snowfall in the Southern hemisphere**. To make things worse, crocuses' spatial properties can be defined by a turtle algorithm.[56] This algorithm can produce a pixel pattern on a computer screen (that happens to look just like a crocus). So there is a translation resemblance between the screen pattern

---

[55] Quoted from Fred Adams, et. al., "The Rutgers Chainsaw Massacre IV: Jason Goes Sailing". Unpublished draft.

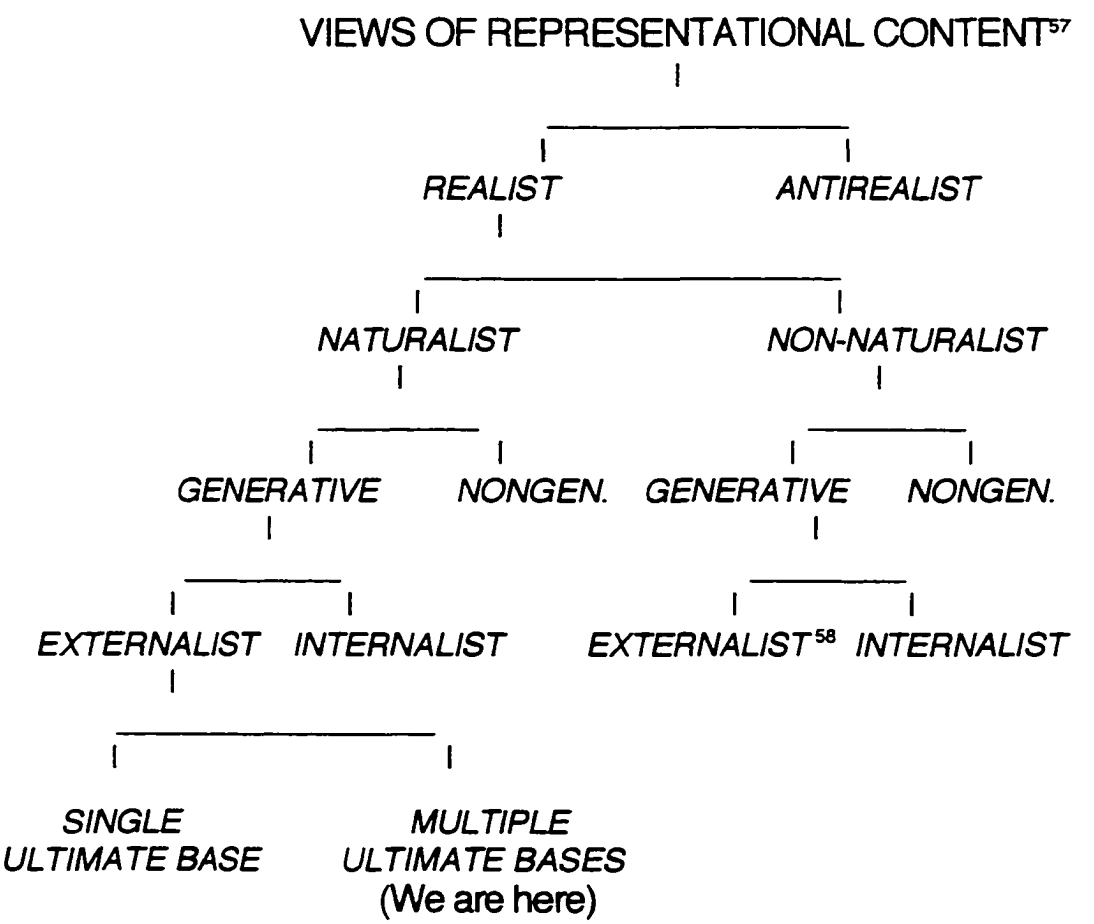
[56] "Turtle algorithms" are mathematical recursions that when allowed to cycle produce remarkable images of real-life objects. Bontanical-like examples can be found in Martin Schroeder, *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise.* (New York: W. H. Freeman and Company, 1991), pp. 259-261 and plates 7, a-f.

and the visual cortex image--the one at work in the turtle algorithm. So the visual cortex process also has the content **such and such pixel pattern**. What could be more absurd?

At first glance multiple contents can be counter-intuitive. It seems to many that we know our contents are singular in an introspective way. This can be the prime motive for an explanation of content that focuses, ultimately, on only one generator. Other basic worries include these: Will allowing all these content generators create too many contents for us to navigate? Will it require we build a complex theoretical apparatus which allows only one generator to act at different times or which somehow combine their contents coherently? My view is "none of the above". What I hope to show in coming chapters is that multiple contents like the one above are *not* absurd; they can be very useful for both organisms and within our theory of content. I also hope to undermine the *intuition* that our contents are singular.

Summary, Chapters 1 and 2

We can help put the situation into context with this diagram:



We are pursuing the following view: A realist, naturalist, generative externalism that admits many ultimate bases for the generation of content. Further, we will not attempt to resolve the

[57] John Haldane offers a portion of this diagram in his, "Putnam and Intentionality", *Philosophy and Phenomenological Research*, Vol. LII, No. 3 (1992), along with the positions of some contemporary authors within it.

[58] The single/multiple option is also available for nonnaturalist externalists. Space doesn't permit its display.

choice problem, but will instead develop an account that takes advantage of the existence of multiple contents associated with a single representational token.

Chapter III. Process/Relationship Relativity and Multiple Contents

Content and Naturalism, the new direction of process/relationship relativity

In this chapter I'll draw on a number of points to argue for a new direction, the *process and relationship relativity of content,* "process relativity" for short. Supporting this, in the next chapter I will also argue that none of the current popular theories of representational content can solve the profoundly central *problem of misrepresentation.*The problem of misrepresentation is the problem of showing how our representations can be *false.* Nothing real except our representational contents can *be* something that is, to speak paradoxically, *nonexistent.* Misrepresentation reflects this mysterious autonomy our contents have from brute reality. Because of this mass failure we must look for a new direction from naturalism--the process relativity of content is a compelling one. But before we field process/relationship relativity against misrepresentation and other central problems, let's first make clear what the idea is and why it's a very reasonable direction for a Naturalist to take.

What is the process and relationship relativity of content?

The thesis of process/relationship relativity is simple. It is a natural consequence of an ecumenical attitude about content generators. It involves three related claims,

1) Representational content C exists wholly relative to a physical process or pattern of relations that the bearer of that content (the "representational token") is part of. No process or pattern of relationships, no content.

2) The physical parts and their physical relations in this process, or included in the pattern of relations, fix what the content of C is.

3) Different representational contents can be had by the same representational token at the same time when this token is part of different content-generating physical processes or relationships.

4) As developed here, the processes and relationships mentioned above are typically *external* to the representing system in some respects.

So the hypothesis is that representational content is a matter of being part of a certain kind of physical process or being embedded in certain patterns of physical relations. Different processes and physical relations provide different content. As we discriminate these, we discriminate contents.

(1) and (2) are distinct claims. (1) says content requires a physical process or relationship to exist. (2) extends (1) to the fixing of content. A physical process or pattern of physical relations is sufficient for the existence of each particular content. (3) is entailed by (2) when we add the observation that different physical processes and relations can generate content and these can simultaneously impinge upon the same representational token. This can prove very useful in actually deploying process/relationship relativity against some sticky problems facing all theories of content. Taken together, (1), (2) and (3) give us the thesis of the process/relationship relativity of content. In our development of process/relationship relativity is a fourth thesis,*externalism* about external world contents. We have already defended this. Let's see some reasons for the parts 1-3 of the thesis next.

A Naturalist argument for process and relationship relativity

This argument, via Naturalism, is a good starting place for

exploring the view,

1) Naturalism is true.

2) If (1), then the only things that exist are Natural entities or derived from these; physical individuals, properties and their patterns of change (processes, individual and type) and the properties and relations that are generated from the combination of these.

3) Content properties, while real, are not basic physical properties the way fractional charge, rest mass or kinetic energy might be.

4) So (given 1, 2 and 3), there must be a mechanism of their generation from more basic physical properties.

5) The only reasonable candidate for the generation of content properties is they are produced by Natural processes and patterns of relationships. Content is not a substance composition like curry powder, it is not like an object or monadic property that exists in the absence of a larger process or pattern of relationships.

6) So representational content exists within a process or pattern of physical relations. It is a *process or relation embedded* entity.

7) Because these contents exist within physical processes and patterns, and representationally differ in accordance with their associated processes and patterns, the natural explanation of their content is that it is generated by their representational tokens' position in these differing processes (individual and/or type of process or relationship).

This is an argument for the "process/relationship

relativity" of content. But we shouldn't stop here. I would add this,

> 8) Given (6) and (7), the same representational token may have many distinct contents, because any representational token can simultaneously be embedded in any number of content generating processes and relationships and there is no resource within the Naturalist ontology for elevating some of these putative contents above others as *the* content.
>
> 9) So Naturalism implies the existence of multiple contents for some (if not all) representational tokens.

This is not a deductive argument for process/relationship relativity. It is a deductive argument for the conclusion that process/relationship relativity is a reasonable, probable inference from Naturalism. In the terminology of the Introduction, process/relationship relativity provides an obvious explanatory opportunity if our starting point is Naturalism. Premise (1) is stipulated. Premise (2) simply elaborates (1). (3) seems obvious. (4) strikes me as reasonable.[59] It's steps (5)-(9) that require discussion.

Contents are process/relationship embedded entities

Premise (5) is important and straightforward. The idea is that it is only in the context of larger processes and relationships

[59] I have argued for it at length elsewhere.

that content exists. This is part (1) of the process/relationship relativity thesis. These processes include interactions between a cognitive system and its environment as well as processes within the cognitive system itself. This conforms to our actual experience perfectly: Try to imagine a representation that is not dependent on some larger scenario of events. *None* of our representations enjoy this status. All are the result of processes before or after our birth--childhood and subsequent development and perhaps evolutionary factors too. Content, as we find it, is process and relation embedded. We've seen this is also the direction we find our best explanatory opportunity lies in.

Premise (6) offers the basic hypothesis that physical processes and relations generate content. (7) develops this by claiming that the nature of these processes and relations determines which contents are generated in any particular case. This is the most straightforward move to make: What is generating x determines the form of x. Here, as everywhere else in this topic, concerns about *explanatory opportunity* loom large. The processes and relations any representational token is caught up in are complex and varied. But immediately we notice (8), and are probably surprised.....

Why should we expect multiple contents?

Premise (8) is as critical as it is unexpected. It has something of a pedigree. We can see an open suggestion of (8) in *the depth problem* .[60] The depth problem appears for theories of perceptual content that make claims like this one,

> R has the content **C** when R's activation reliably indicates (*or* is caused by) C.

Suppose C is my feline friend, Little Cat B. I see him and R activates. Then R represents Little Cat B. But R's activation also indicates (and is caused by) the image on my retina, the neural depolarizations in my optic nerve, the neural prepossessing in the visual cortex and so on. Why is it R's content not **retinal activation A**, **depolarizations B**, or so on? All are equally implied by the claim of our perceptual theory. Why is it external world objects that our perceptions represent?

The best attempt to solve the depth problem that I know of is Dretske's.[61] At first glance his solution looks reasonable, obvious even. He argues that because our perceptions keep certain features of external world objects constant, they represent these objects. When we shake our heads objects do

---

[60] See Sterelny's *The Representational Theory of Mind: An Introduction* (Cambridge:Basil Blackwell, 1991) for an clear discussion of the depth problem.

[61] Fred Dretske,*Knowledge and the Flow of Information*. Cambridge: MIT Press, 1981.

not appear to move (much). But the image projected on our retina moves a great deal. Objects also appear to be the same size at different distances, and the same color under different illuminations. This shows that our perceptual content is fixed at only one "depth"--the object depth.

But once critically considered, this constancy is both dubious and beside the point. It's dubious because an equally available interpretation shows a relentless *inconstancy*. This is the inconstancy so celebrated by Idealists. An octagonal tower appears round in the distance. For a certain strain of Idealist this means that it *really is* round when "at a distance". As for size constancy, the fact is that we are just as able to judge that objects do *look* smaller when farther away or *look* to be moving when shaking the head as we are that they don't *look* smaller or in motion. If I judge size like a painter, hold my thumb at arm's length up to everything, objects shrink with distance. If I attend to my nose and eyebrows while I shake my head, relative to these reference points I can judge objects to be moving. And it seems manifest that objects *do* change their color in different lighting. These judgments are as easily invoked as our ordinary ones. We often assume that this doesn't reflect object changes, but this can be easily overcome by other evidence that an object is indeed shrinking or moving or changing hue. The fact

that we can easily entertain all these contrary appearances demonstrates that constancy judgments are *not given* by perception, but involve an additional judgment, one that *presupposes* that our perceptual experiences are of stable objects instead of demonstrating they are.

Dretske's solution is also beside the point because the critic can simply reply that the object-interpretation of content is only more constant than others--not that this interpretation is correct. There is no obvious reason to link constancy with *the only actual content.* Why can't contents be inconstant? There is no *a priori* reason. An Idealism where color-patch perceptions ("objects") really do shrink with increasing perceptions of distance is not *a priori* false, even though constancy is better served with a materialistic, object-representing interpretation. This is an especially relevant point if we believe all the contents suggested by the depth problem are equally present. Further, even if we grant that it is to the greatest constancy that the prize of being *the content* goes, recall that our typical size and stillness judgments are neurologically produced. So the depth problem simply reappears with Dretske's observation of constancy. For there is an equal or better constancy in seeing *these neural processes at work in our perceptual judgments.* Why not say that what is being held constant when I see a

nonshrinking train move away are these neurological processes' effect on my judgments and accompanying assertions, instead of the train?

So Dretske's solution fails both in its claim of constancy and the relevance of this claim. The process/relationship relativity diagnosis of this failure is simple: The depth problem has no solution. It is a version of a more general and misguided aim of most representational theorizing, the search for a solution to the "choice problem"--the fixing of a single content for each representational token. We will return to this at length in the following chapters. For now notice that our preoccupation with finding a solution to the depth problem which fixes objects as our perceptual contents appears to be an artifact of our particular situation in the world, not of any deep considerations concerning the content of our visual cortex states. Imagine humans living as a community of neuro-experimentalists--fanatical scientists interested in little but researching*neural processes*. Their existence is almost entirely mediated by their instruments, with sufficient life-dependence on machines (digital/biological interfaces and the rest of the paraphernalia envisioned by current science fiction) so that they rarely if ever have to interact with their environment via biological senses. It would be quite "obvious" to them that included in the

representational significance of their visual (and other) cortex activities is the indication of neural structures/events, not simply external world objects. This "obviousness" would stem from two sources; one is that this is what these activities *indicate with a very high degree of probability*, the other is that this indication is *congruent with the activities of the community*--theorizing about the brain and how its systems inter-relate. (We might call these neuro-experimentalists *fanatic internalists.*)

We can also imagine (and perhaps have met, on occasion) the fanatic *neurophilosophers.* These individuals claim to see, not just people and shirts and food with their eyes, but at the very same time waves of neural depolarization, vector generation and so on in the visual cortex. Though this is not the standard way of reporting visual experiences, it is accurate. They are witnessing exactly those things. At this very moment, so are we.[62]

But we are not such a community (for the most part). *Our* preoccupation is with navigating through an environment where biological senses and their relations to the external world are critical. So our preoccupation with the external-world contents of the perceptual system follows the same rationale as the

[62] Paul Churchland once launched into such a series of utterances at dinner. Though it struck me as bizarre at the time, he was correct. The mind is in some ways manifest to itself. In our case, what is manifest is actual brain processes. What is *not manifest* is that these are brain processes. But that's a problem with external objects, too. My son once tried, with great exertion, to eat an orange tennis ball....

neuro-experimentalists--indication and relevance to our activities. Until we attempt neuroscience, it matters not at all what the indication relations between neural events are. Stepping back from both these communities I would have to agree with the neuro-experimentalists *and* us. The point is that there are any number of representational contents we might find within the same representational token, all of which are equally correct. Moral of the story: Attention to some contents and not others is an entirely different phenomenon than content itself.

We can illustrate this even more clearly if we envision cognitive systems with *communal elements.* Here's one: In humankind long term memory is stored in distributed areas of the brain and no human shares memory tissue with another. But imagine a compartmental and shared memory bank, one that is used by a large number of individuals. Suppose this bank is "hard-wired" with several trillion episode types. It is a sort of archive of possible memory types, amassed over thousands of years through the combined input of billions of cognizers. An event is stored by a particular cognizer in a short term memory. Then using various dimensions of resemblance the short term memory is assigned to one of the trillions of permanent memories. If the resemblance is too weak (given a stipulated

threshold) for assignment to any of these, the event is added as a new type. (At the very worst, this assignment is like a minutely degraded copy of the short term memory--which is surely the situation with human long term memories too.) Then the cognizer is given a memory address for this representational token. When the cognizer seeks the memory of this event, it accesses this location in the bank. Naturally, at this point in the bank's history few events are ever added. Almost all are assigned to old locations.

Two points follow from our community memory bank story. (1) representational tokens can have different memories assigned to them--the original and all the later ones, and more interestingly, (2) that this can happen at the *same time*. Different individuals can simultaneously have different events assigned to the same memory token in the bank. This is an example of a single representational token simultaneously having multiple contents, each relative to different processes it is embedded in, here different short term memory events from different individuals being mapped onto the same token.

These and any number of other examples make a case for multiple contents. But none of them harvest the stronger, simple argument naturalism and process/relationship relativity allows.

A. Naturalism and Escaping Multiple Contents

If we accept that processes and physical relations generate content and notice that many independent processes and patterns of relations are at work on any representational token, each of which appear to be content generating, we have multiple contents. *If we are to escape multiple contents we need a way to act on this set so just one content is generated.*

There are two ways we might proceed. First, we might see if the many contents enjoyed by each token can be combined into a "master content" that involves a single coherent content. Of course combining different contents assumes there are different ones to combine, so we haven't gotten out of multiple contents, we just interrelate them for an additional, single content. But the real problem is that there is no way this could be a *coherent* content. It would be a mangled hodgepodge of very different, often contradictory features, a monster of no use in this world. To keep things simple, just consider the contents the depth problem would have us combine: **tree**, **neural activation cascade x** and **quantum waveform z**. Some things are simply not meant to be.

The second route is more promising: *Screen all but one of these contents out.* But which ones and how?Again, this is what

I call the *choice problem: How do we arrive at a single content for each representational token?* In the following chapters we will explore the choice problem in various contexts, but for now let's just focus on how a *screening property* might be sought.

The token itself cannot do the trick because it is the focus of multiple processes and relations, all reasonably content generating. We need to look beyond the token to generate a screening property. Could it be the token's relations to other tokens? These relations are either contentful in nature or not. By "contentful in nature" I mean facts that consist of the larger pattern of contents the system has.

One example of contentful relations are content to content relations like "is" and "has"of *semantic network theory*-- a **robin** is a **bird** and has **wings**. By itself this does nothing. Because contents are multiple for each token, all we get is: **C** "is" **B**, **C**' "is" **B**' and so on.

Perhaps it is the overall contentful *situation* of the tokens that has screening powers. Of course asking *how* is important, but let's ignore the worry about how exactly this is going to work (I have no idea) and focus on a more basic problem. We can understand this "situation" as consisting of all the multiple contents of all the tokens or some subset of these. The subset version only recreates our original issue: The problem with this

direction is specifying *which* contentful situation. Because contents are multiple, so are the subsets of contentful "situations". It's the choice problem all over again. So suppose, instead, we try to build a single situation out of all the contents had by each token. But then the problem is that we have already based the situation on the acceptance of all of these contents. The contents are already present to produce the broader content situation. We have embraced, not screened, the plethora of contents.[63]

Noncontentful relationships the token bears to other entities remain to do the screening. The essential question again: How might these effect the screening of all but one content? There are two ways we can conceive the situation. The first is that the contents are had by the token and all but one are *causally removed* by noncontentful factors. The second is that the content generating relationships that otherwise would generate multiple contents are somehow prevented for acting on the token, with the exception of *one*.

The first option looks hopeless. Contents are not substances, they are not objects. They are not the sort of things that once present can then be extracted *at a certain point in*

[63] This is similar to the problem semantic network theory has in explaining the content of concepts like "bird". It's true that concepts can be understood as occupying a position in a network of other concepts, but this doesn't let us account for the fact that any of the concepts in such a network have content. This is required at the beginning of the story.

time. If the relationships are in place, they generate content. This effect cannot be "removed" after the fact. Notice this point applies equally to the contentful based approaches to screening mentioned above.

The second option is less bizarre but equally mysterious. Relationships exist and if they are content generating then they do exactly that--generate content. This means that if we are to indulge in content screening then the relationships in question--say, causal--must be rendered inert in their content-generating powers. Our situation seems hopeless: We create the equivalent of a world where only the token and one content generating relationship to the environment exist, the simplest case of content generation. The trouble here is that in the simplest case the relationships in question *do* generate content. Adding other properties and relationships to the simplest case either (1) has no effect on the existence of this case in a more complex context or somehow (2)*removes* the simplest case as part of the correct description of the token's relationships with the world. (1) does nothing to screen contents, (2) reduces to the first option; extraction. But again, this notion of *extracting a preexisting content* is mysterious. What we are trying to do is really no different than putting X to the right of Y, and then trying to make X stop being to the right of Y by painting it. I have no

idea what it would be to remove a content without removing the process or relationship that generates it--causality, indication....

Psychological ecology and alternative theories

On psychological ecology tokens have multiple contents. Process and relationship relativity embraces this and as we'll see, uses it to solve a number critical problems for any theory of content. But we can imagine alternative theories that begin with psychological ecology as a solution to the source problem, then use explanatory-relevance standards to exclude some contents and not others and isolating one content for each token. Such theories then proceed to explain behavior and cognition with this limited subset. I believe this is the best way to understand Dretske's solution to the depth problem. He wants contents to be used in explaining behavior--predator avoidance behavior, foraging behavior and so on. So he introduces a standard to help pick out those contents relevant to this explanatory aim.

We might ask whether this is inconsistent with the process/relationship relativity approach to psychological ecology. I believe the answer is *no*. If our theoretical method begins with psychological ecology and introduces a module of exclusion and then proceeds to explain, the issue is not whether there really is one content for each token--there isn't--but

whether a theory that focuses on only one content has theoretical and explanatory advantages over process/relationship relativity in its explanatory domain. Much like alternative conceptions of quantum mechanics, alternative theories built on psychological ecology can be embraced without rejecting its answer to the source problem, one that leaves us with multiple contents at the beginning.

B. Conclusion: No Escape from Multiple Contents, None Necessary

The essential problem is that we cannot *get rid* of the unwanted contents, we can only hope for a way to *ignore some and not others* in our preferred content ascriptions. I believe this is what is really being gotten at by the one token, one content assumption: One token, one content *ascription.* This is all well and fine for multiple contents and reveals an easy oversight to make; the issue of what content assignments we make is distinct from the issue of what contents are present in a system. There are two ways this can happen: (1) Instrumentalism, where we make content assignments without any attempt to determine if they are literally true, or (2) *realist but interest relative assignments.* Process/relationship relativity is a realist

theory of content. The assignments we make within this framework are literally correct, but derive from issues distinct from just what contents are there. Consider the depth problem again. The different assignments a *neurologist* and a *behavioral ecologist* would give to the visual process beginning with a lion are a perfect example.

How do we proceed in a multiple content, interest relative world of content ascription? Here's how: Adopt a theory of explanation with generalities about how contents associate with other contents and behaviors. We all know one of these: folk psychology. Next, show a system probably has certain contents. Then, given the generalities, show that these contents lead to the contents or behavior in question.

I think this is what we do in any case. When we ascribe contents it is with a certain view to what the system is. We must adopt a certain framework of assumptions towards the system prior to ascribing contents and relating these to behavior. A great and complex controversy in recent philosophy of mind focuses on just this fact. Writers like Dennett and Davidson have adopted a principle of *rationality* in their accounts of proper ascription.[64] Others, like Stich, have claimed that there are

[64] See Dennett's *The Intentional Stance* (Cambridge: MIT Press, 1987) chapter 2. See also and Davidson's "Radical Interpretation," *Dialectica*, 27, pp. 313-328.

cases where it is perfectly appropriate to assume irrationality.[65] I would side with the latter. But what is important is that our ascriptions are not only driven by what contents are present but also by our explanatory aims and the framework of assumptions we use to characterize the system. This, again, is perfectly compatible with multiple contents.

The only real oddity in all this is the following possibility: Two very different content-driven explanations for the same behavior by the same system might *both* be true. It could even happen that one pattern of ascription reveals a person to be rational in her behavior and the other reveals her to be possessed in the very same situation by bizarre motives and irrational beliefs. It's not *impossible.* I don't think this is all that surprising. The mind has many "levels" and is rich in convolution. Clinical psychology purchases its existence from this sometimes harrowing, sometimes amusing fact. Multiple contents provide a new and interesting way to understand it. More abstractly, we can imagine a space that represents the range of contents that a system enjoys via its external relations and the processes it involves and is embedded in. There are many equally real trajectories through this space that map onto

[65] Stich launches this critique in "Dennett on Intentional Systems," *Philosophical Topics,* 12, pp. 38-62. Sterelny offers a similar critique in *The Representational Theory of Mind* (Cambridge: Basil Blackwell, 1990), pp. 94-96.

the same behavior. There is nothing about this that strikes me as either surprising or problematic.

A more important issue: Is *disagreement* --as opposed to *difference*--about the contents we ascribe to a system impossible? Is "everyone right"? No. Ascriptions can still be false when the contents in question are not present, when the generalities employed are flawed, when the framework of assumptions about the system's nature is misguided, or any combination of these.

Why is the presence of multiple contents surprising?

If the existence of multiple contents is so reasonable, why do we habitually think and speak as if our representational tokens have only one content? I'm not sure this is the right interpretation of our behavior. We do ascribe single contents, but this doesn't show we are, in our ascription behavior, asserting there are only these contents at work. Indeed, I don't think we are all that reflective about it. But it is a very popular stance the moment we do become reflective. Why? I believe this is an illusion produced by our typical way of describing representational phenomenon. There are two steps in this illusion: We assume the content of a representational token is what we *say* it is and we assume what we say is the end of the

story. The first step is exactly the sort of thing a cognitive scientist would be rightfully dissatisfied with. Many of the properties of our representational abilities are obscure in our verbal reports, and even at odds with them. (Which isn't to say we are not often correct in our reports, too.) Step two is the sort of assumption both a cognitivist and philosopher should be suspicious of. But both are how we are taught to express ourselves and the pedigree of this teaching, like most all our prescientific "psychology", is nothing but pragmatic in kind.

Suppose I truthfully report, "I am seeing a tree". My visual cortex process is the representational token. There is a language cortex process that attends it, one that produces (along with certain frontal cortex processes) the announcement, "I am seeing a tree." The representational token in the visual system has been mapped onto a single entity, my sentence. My description of its content ends at this point. But need it? On the assumption that we have *fully described the content of the visual process* it looks like there is one content present. But as easy to make as this assumption is, we have to ask, is this true? Why think so? Recall our critique of Dretske's solution to the depth problem. True, the sentence is *not* , "I see a retinal activation pattern of type n." But what if it was? Why say that a mistake has been made? Our neuro-

experimentalists wouldn't. If we were considering a long term memory of this visual cortex process, our communal memory users needn't. It is just this assumption of *complete* verbal description that process/relationship relativity challenges.

An alternative and more plausible explanation is readily available: We *tend* to assign one sentence to a representational token (though our stories show we needn't) because in most contexts it is useful to make only one assignment. Why do we make the single assignments we do?This would best be governed by *relevance criteria*. I imagine it is, not by any absolute fact about R having just one content.

Can we use a generative theory to show that multiple contents don't exist?

Our discussion has been based on extemalism. In the face of this broad based account, the unrepentant internalist (or uncomfortable externalist) may consider the abstract possibility of *demonstrating* that contents are singular--an *a priori* counter strike, via an alternative generative theory.[66] I believe fundamental doubts surround this project.

There are two perspectives one might take on the ultimate task of a generative theory of content: (1) we know

[66] Here I contrast demonstrative proof with probable or methodological considerations.

what the contents of our representational states are and we are trying to give a generative account that reproduces this knowledge from explanatory principles--it gives the how and why of what we already know. Or (2), our knowledge of our representational contents is itself a representational construction that ought, in the end, to be guided by what our theory tells us these contents are, given the assumptions of the theory. Negotiations between these two extremes are possible, but they seem to represent the range from a prior, introspective knowledge-of-content view and an explicit theoretical-construct view of such knowledge.

Both these positions have their complications. I prefer the second. I believe the first is ultimately untenable, but will not dispute it. What I would like to show is that on either of these conceptions there is no way to *rule out a priori* multiple contents. Let's take each in turn.

If we have prior knowledge of the contents of our representational states and deny they possess multiple contents, we need to show that they do not possess contents we don't know about. The fact that we know of some contents and not others doesn't show the others are not present--this would be to argue from ignorance to a positive conclusion. So our immediate impressions cannot settle the issue. That means

we must turn to a generative theory to show that only particular contents--the ones we know of prior to theorizing--are generated. What happens when we make the attempt? We fall into a circle of reasoning. Here's a general observation I believe is critical to understanding our dilemma,

> Any generative theory of content is itself a representational system of contents.This theory cannot provide an account of the content of representational tokens without our first assuming its terms and semantic relations have certain contents. Its interpretation and subsequent entailments are based on this assumed interpretation. The theory itself cannot supply the content assignment to its own terms because it requires these already be in place to entail anything.

The problem I see is this: The fact that a generative theory entails singular contents for representational tokens doesn't show us that *in fact* their contents are singular, because the terms of the theory may themselves *in fact* have multiple contents, and the theory, may not recognize this. Generative theories of content are self-blind to the content assignments of their own terms, on pain of circularity. Indeed, there is the possibility that for any generative theory that only assigns single contents, the correct content assignment to its terms is multiple, and among these multiple contents is a set under which the theory actually entails *multiple contents for any representation.* The theory itself can't tell if its terms fall under this scenario or

not, because it cannot fix the contents, or the singularity or multiplicity of contents, of its own terms. Put a slightly different way: The advocate of singular contents can't be sure she is really dealing with one theory--her theory may in fact involve a set of different theories, each one corresponding to a different content-interpretation of its terms.

This point immediately extends to the latter option; that our theory itself determines our content assignments. We have to weather a certain humbleness here. We must recognize that our account inherently leaves open the possibility of multiple contents, no matter what its entailments, so that under different content assignments, we get different entailments. But notice that the above reasoning doesn't go the other way, equally undermining a generative theory's entailment of multiple contents: If our theory entails multiple contents, the fact that its terms have multiple contents doesn't show that there is an interpretation where *our theory* entails singular contents--because that would be a different theory than the one we hold--the same physical manifestation (pattern of tokens, be these linguistic or neurological) under a different content assignment. *The problem the advocate of singular content labors under is that her favored generative theory might always involve, in fact, a different theory than she realizes--any number of different*

theories--since she believes it has only a single pattern of content assignment to its terms when it actually may have many.

For this reason I don't believe we can definitively demonstrate via a generative theory that our representational tokens have only one content.[67]

Conclusion: Summing up and the road ahead

We cannot escape the simple fact that on externalism representational tokens are embedded in multiple processes and relations, many of which are valid candidates for content generation. Nor can we, I think, escape this: *Whatever powers might be loose in the Natural world, there is no superior level of reality to bestow special privilege on some of these and not others.* They appear all on a par, equally generating. All fulfill the function of generating the *syndrome of aboutness* outlined in chapter 1.

Here is where metaphysics makes its difference. If we were Platonist's (in the traditional two-worlds sense) there would be a privileged standard--"participation" in the forms, that might be appealed to in order to fix the contents of our

[67] Indeed, if we feel like playing the odds, perhaps the odds are against the idea of singular content. After all, there are any number kinds of content scenarios our representational tokens *might* enjoy. All but one of these kinds is a singular content one. The others have this form: 1+n contents. "n" could be any value.

representations.[68] They are *formed* by the entities of a higher reality--the Forms themselves. But on Naturalism no such reality is available, and we sink back into the swarm of content generating processes and relations every representational token is surrounded by.

It is only under the pressure of the choice problem that we have any reason to set some of these generators aside in favor of one. Any number of physical relationships to a cognitive system reasonably generate content; world to system causation, indication and resemblance all appear to by their very nature. But the choice problem needn't be a problem once we distinguish content ascription from content possession.

We will see the challenge to the *one representational token/one content* assumption continued in the next chapter when we discuss current theories of misrepresentation and their short comings. After rejecting these attempts, we will see how profitable giving up this assumption can be, via process/pattern relativity's approach to misrepresentation and the other classic problems of content.

---

[68] I assume a standard reading of Plato's theory of Forms as presented in *The Republic*, Books VI and VII.

Chapter IV. The Problem of Misrepresentation

The problem defined

The problem of misrepresentation is one of the most critical hurdles any theory of representation must pass. It highlights two features of representation that define the whole phenomenon, *aboutness* (say, aboutness concerning the world) and the power of aboutness to be in some sense autonomous from the world. Our representations can *fail to be true*. We can *mis*represent our world. "The world hath another to attend it", and this other, representational world is often rebellious in its content. Only through great efforts and mechanizations does the world of representation match the external world. Usually, in one way or another, it does not. But how is this possible? The answer is far more elusive than might first be thought.

Relative to different theories of content, the problem appears in different ways. Let's review some examples of otherwise likely theories failing to solve the problem, distill out the shared aspect of their failure, and diagnose the source of their shared failing.

First highlighted in recent literature by Dretske, the problem of misrepresentation has received a great deal of

attention.[69] Three approaches are prominent in the naturalist tradition, Fodor's appeal to "asymmetrical dependence", Dretske's indication/teleological hybrid and Papineau's and Millikan's teleological theories drawn from evolutionary biology. In addition to these standard contemporary views, an appealing newcomer can be found in Austen Clark's information increase approach. [70]

These theories are leading contenders in the literature, all receiving extensive discussion, with the exception of Clark's more recent addition. That in itself is one reason to focus on them--we are looking at the defining mainstream. *But the main reason is to reveal a thematic problem with such accounts--their focus on the choice problem as the central problem to be solved.* We'll critique each of these by looking both at their viability as generative explanations and as logically adequate solutions. Because our project is to sketch a generative theory of representation, the former task is decisive. But more general criticisms can be instructive too, not just because they are valid

---

[69] See his classic, *Knowledge and the Flow of Information* (Cambridge:MIT Press, 1981), ch. 8.

[70] The best exposition of Fodor's theory is found in his *A Theory of Content and Other Essays* (Cambridge:MIT Press, 1990) pp. 89-136. David Papineau explores a simple teleological approach in, "Representation and Explanation."*Philosophy of Science* 5, pp. 550-72. Ruth Millikan launches her more complex view in *Language, Thought and Other Biological Categories* (Cambridge:MIT Press, 1984). See also, "Biosemantics" *Journal of Philosophy* 86, pp. 281-97.
Austen Clark's approach is found in his, "Mice, Shrews and Misrepresentation." *Journal of Philosophy, XX* (1993), pp. 290-310.

from the naturalist perspective but because they also make us sensitive to the requirements of an overall satisfactory approach to error.

The choice problem again

The traditional focus of the problem of misrepresentation is the *choice problem*. Of all the possible contents a representation might have, how do we choose one as *the* content of the representation, so that this content can be tokened when, in fact, it is an *error* to do so? In the context of a causal theory of content, we try to isolate one of these causal source-objects (or properties) as providing *the* content, so that when other objects or properties activate the token, it's an error. Hence the choice problem for a causal theory: How do we justify the choice of one activating object as giving the correct content out of all the other activating objects or properties?

In this chapter the basic challenge for all these theories will be to solve the *choice problem*. But there are other important challenges to tackle too, as we'll see.

A. Contemporary Approaches to Misrepresentation

A solution to the misrepresentation problem is always

going to be relative to a theory of content generation. We'll give (very) modest reviews of these theories as we take a look at the solutions.

1. Indicator theories; Fodor and asymmetrical dependence

Fodor offers a version of indication semantics ("covariance" or "information content") in his book, *Psychosemantics*.[71] He expands upon these ideas in *A Theory of Content and Other Essays,* exploring the concept of *"robustness".*[72] Robustness is the independence of the content of a representation from the ways it is activated. What is "robustness" in the final analysis? The recognition of the choice *problem.*

As with all indication theories the initial idea is simple,

A tokening of a representation R is a representation of the type F if it is correlated with the presence of Fs within the sensory range of the cognitive system.

The choice problem appears immediately: If cows activate R reliably, but so do buffalo, R's content becomes a disjunction of the different entities that activate it. Error becomes impossible, because some member of the disjunction is always

[71] Cambridge: MIT Press, 1987.
[72] See chapter 2.

present. [73]

But notice that any correlation between real buffalo and cow representations is dependent on a correlation between the presence of cows (the property of being a cow) and cow representations but *not vice versa*. The correlation between cows and cow representations would exist even if there were no buffalo. The dependence doesn't go both ways, it is *asymmetrical.*

Here Fodor detects a solution for the problem of misrepresentation and finds an opportunity to imbue representations with "robustness". In real disjunctive representations, there is no asymmetrical dependence. In **PvQ** the connections between P and representation R are on a par with the connections between Q and R. So the representation's content is **PvQ**. This isn't the case with the connection between cows and **cow** and buffalo and **cow**. The buffalo-**cow** connection to R would not exist if the cow-**cow** connection did not. So R represents cows, not cows-or-buffalo. This

[73] An aside: F doesn't have to cause R. Indication doesn't have to be causal. Mind/body parallelism theories of 18th century dualism are compatible with indication semantics. The things/types in the world were said to have no causal effect on the indicating representations with the mind. God arranged the pattern of representation activation motor response, etc. to behave as if there were causal links to the world. No world-mind causation but the mind still rife with indication, divinely inspired no less.

But given contemporary metaphysics the correlation is established via an external world type causing the tokening of the representation. While I believe indication needn't be causal in even a wholly physical, godless world, and still generate useful and true contents but this isn't important here. So here misrepresentation problem is just the problem described for causal theories, in the form of the disjunction problem.

asymmetric dependence resolves into a pair of counterfactuals. It's true,

a) If cows didn't produce tokenings of cow representations buffalo would not produce these tokenings.

while it is not true that,

b) If buffalo didn't produce tokenings of cow representations cows would not produce these tokenings.

So the representation-tokening dependence runs only one way, from cow representations caused by cows to cow representations caused by buffalo. So Fodor's claim is simple: Find this one way dependence and you find a principled reason to say that the content isn't **cow or buffalo** but **cow**. Buffalo are sometimes mistaken for cows.

In summary, the theory is this,

1) The property of being a cow has a law-like relation to the tokening of **cow** in a cognitive system like the human brain. This connects us to the world and is part of the process that generates contents like **cow**.
2) There are in fact some **cow** tokenings caused by the

property of being a cow.

3) For anything that doesn't have the property of being a cow (buffalo, for instance), if it causes **cow**, then its causing **cow** is asymmetrically dependent on the property of being a cow causing **cow**. This allows **cow** to be the correct, singular representational content.

4) There are actually or potentially some not-cow caused tokenings of **cow**. Fodor calls this possibility *robustness*. The point of this is to allow for not just for cases of misrepresentation, but also for the tokening of **cow** in scenarios like questions ("Where does the milk in cheddar cheese come from?"), and associations, like the sight of a diary farm.

Fodor thinks he is on to more than a solution to the disjunction problem. He has tackled the essential problem of producing content in a wholly natural world: He has a theory of content on his hands.

A critique of asymmetrical dependence

This is an engaging idea. Many worries follow in its wake though. These are drawn from scenarios of misrepresentation that occur where there is nothing of the asymmetry, the ontology of counter factuals, the demands of adequate

explanation and the assumption of a solution to the choice problem.

Error in cases where there is no asymmetry of dependence

First let's play the counter-example game. Counter examples are useful at testing a theory. The downside is that they don't often explicitly pinpoint *where* the theory goes wrong. I'll suggest a diagnosis afterwards.

Even if we grant that Fodor's counterfactuals are the kind of thing that could count as an adequate explanation, some authors have questioned that they really do this by imagining scenarios where *both* (a) and (b) are true. In these cases the one way dependency breaks down, but the system still misrepresents. It might be that these scenarios never happened and never will, but that's no matter. They are objections to the claim that Fodor's theory captures what is *universally sufficient* to the idea of misrepresentation--what is true in "all relevant possible worlds" where misrepresentation happens. So these are important if one demands understanding that goes beyond certain coincidental patterns in misrepresentation in our particular corner of reality.

Cummins imagines a scenario were children are taught to catch mice for a tribal potion. Mice are rare but shrews are

common. The children learn **mouse** by practice catching shrews. They never see mice (perhaps a few too many potions and now mice are extinct) and they wouldn't recognize a mouse as a mouse (they'd think it was a shrew), but they know shrews are not mice, because they know only mice will work in the potion. The problem for asymmetry theory is this: While the connection between shrews and mice wouldn't exist without the connection between mice and **mouse**, given the way the children learn **mouse**, the connection between mice and **mouse** also wouldn't exist without the connection between shrews and **mouse**. There is no asymmetry of dependence anymore.[74]

Sterelny offers a similar point. There are possible worlds in which twin cows, robot cows, cow holograms are so consistently bovine that we invariably think "cow" on beholding them. But cows in those worlds are so uncow-like that we never think of them as cows. Asymmetrical dependence solves the disjunction problem only if these counterfactuals can be *nonarbitrarily* excluded.[75]

Sterelny is skeptical that this can be done. He should be. A rather esoteric body theory would probably be required. But

[74] Robert Cummins, *Meaning and Mental Representation* (Cambridge: MIT Press, 1987). p. 79.
[75] Kim Sterelny, *The Representational Theory of Mind: An Introduction* (Cambridge: Basil Blackwell, 1991) p. 121.

even if such a theory can be found we have to ask what sort of road we're going down. We're off to see the wizard of naturalized representation. But there's something strange in naturalizing representational properties like error by building a theory of admissible possible worlds to fix relevance-constraints on counterfactuals for thought-experiments in the theory of representation. How exactly this is grounded in natural science or a reflection of a wholly natural world? If representation--like the representation at work in insects, birds, cats, apes and humans--is a biological capacity then it's surprising to find ourselves so far from the empirical facts in our search misrepresentation's foundation.

Godfrey-Smith has argued that there is no reason to suppose cows have an independent link to the concept of **cow**. We can describe another link between cows and a unnatural kind, call it **K**. If **K** contains some buffalo, then there is no asymmetry of dependence. Both cows and buffalo fall under **K** equally. Fodor's task is to exclude this unnatural kind and an infinite number like it. It doesn't seem the unnaturalness of it will help--many unnatural kinds exist--weed, for instance.[76]

In a different vein, Austen Clark proposes this problem: Because an air/water boundary refracts light, the apparent

[76] Ibid., p. 122.

position of any fish underwater, when seen from above, is always shifted from its true position. The laws of optics insure this; so this shift is law-like. It is also a misrepresentation because anyone throwing a spear at this apparent position will always miss the fish. But there seems to be no veridical representation of the fish position which this is asymmetrically dependent on--the misrepresentation is there from the start.[77]

Finally, in a critique by Fred Adams and Kenneth Aizawa, two strong objections are raised.[78] First, it appears that it is humanly impossible to satisfy Fodor's third condition,

> 3) For anything that doesn't have the property of being a cow (buffalo, for instance), if it causes **cow**, then its causing **cow** is asymmetrically dependent on the property of being a cow causing **cow**.

Here's why: Suppose that it is a law that cows cause **cow** to be tokened in a cognitive system (condition 1) and that at some point a cow actually does cause **cow** in this system (condition 2). Allow that non-cows also cause **cow** in the system--the sight of a farm, for instance (condition 4). All we need to add is condition 3--but can we? The problem is that there are all sorts of "pathological" or "accidental" causes of

[77] Austen Clark, "Mice, Shrews and Misrepresentation." *Journal of Philosophy*, 1993, p. 293.

[78] Fred Adams and Kenneth Aizawa, " 'X' means X: Semantics Fodor-Style." *Minds and Machines* **2**, 1992, pp. 175-183.

cow that are *not* dependent on cows. A severe blow to the head, a brain tumor, the ingestion of mind altering drugs, the occurrence of "random" neural discharges and seizures, all of these might cause **cow** to token. Yet none of these are dependent upon the fact that cows cause **cow**, the way we might think a buffalo in the night causing **cow** might be. Because humans--indeed all complex natural cognitive systems--are not immune to such atypical causal influences, condition 3 is not the sort of thing we can aspire to meet.[79]

Another objection offered by Adams and Aizawa takes advantage of the "twin Earth" cases. Again, twin Earth is a duplicate of Earth in every way except one; "water" on twin Earth is not H_2O, but another indistinguishable chemical, "XYZ". Your twin on twin Earth has never causally interacted with H_2O, just XYZ. So her tokening "water" represents XYZ, not H_2O, just as an Earthling's "water" represents H_2O, not XYZ. If an Earthling with the token "water" was transported to twin Earth, she would be mistaken to token "water" in the presence of XYZ. Condition 1 of Fodor's approach would then have it as a law that on twin Earth XYZ causes "water" tokens and on Earth H_2O causes "water" tokens. But because XYZ and H_2O are in fact indistinguishable, these laws also apply in *both* worlds. It is a law

[79] ibid., p. 180.

on twin Earth that H_2O causes "water" tokens and it is a law that on Earth XYZ causes "water" tokens. This threatens the ability of condition 3 to account for the error made by the Earthling on twin Earth, because there is no asymmetrical dependence between cases where H_2O causes "water" and those where XYZ does. Both are equally laws. If we break the connection between "water" and XYZ on Earth, we must break the connection between "water" and XYZ on twin Earth.[80]

Reprise

Fodor's view has been exposed to a storm of objections. There are many hidden complexities here. Can Fodor deal with these objections? Perhaps. But dealt with or not, there is a deeper concern lurking here, one that occurs to me whenever I consider the *thrust* of these objections. I believe it is essentially this: Though asymmetrical dependence may exist in cases like the cow and buffalo one, it is not the cause, but the result, of a successful solution to finding a ground for robustness, that is, for finding a solution to the choice problem. If we can find these asymmetrical dependencies, then we have found a solution to the disjunction and the choice problem. But how do we account for the fact that they are really there? *Pointing to a*

[80] ibid., pp. 180-181.

consequence of a successful solution to the choice problem, a consequence many have heretofore over looked, is interesting. But it is not itself a solution.

The ontological status of the asymmetrical dependence

One illustration of the above critique is this pattern of reasoning:

Fodor's asymmetrical dependence is realized by two counterfactuals, one true, one false. A generative, Naturalist solution to the problem of misrepresentation is going to be suspicious of an explanatory account based on counterfactuals because the ontology of counterfactuals is unrevealed. What *facts* are being gestured at here? Independent of the representational contents that counterfactuals have, are there *facts* of the form counterfactuals take--facts that can be said to be counterfactual *independent of our so describing them?* The worry is that the answer is, "No.". There are facts, often complicated and even baroque, that are the truth makers for counterfactual statements. But nothing in extra-mental reality corresponds to a counterfactual the way there are corresponding objects for representations like "Greycat" or "up quark", or processes corresponding to the notions of "erosion" or "photon emission". Take this "physical sounding" counterfactual,

If the plutonium atom's nucleus had not been struck by a slow neutron, it would not have fissioned.

The facts that stand behind the statement's truth are not counterfactual facts. There are positive facts of the form, "P is related to Q in way R" and so forth. They include the bonding strength of neutrons to protons, how this structure endures through time and how a neutron passing through this structure causally interacts with the strong nuclear force and splits the nucleus. Explanation in the sense of explaining "how the fissioning was produced" is absent.

The reason that counterfactuals are merely representations is that counterfactuals begin with the assumption something doesn't occur or obtain, and draw consequences based on this hypothetical assumption. That's the "counter" in counterfactuals. But in the world there is not literally, "something such that it doesn't obtain", which produces consequences based simply on its nonexistence. Little could be more obvious. And what if it did exist? That would require Mienong-like ontologies where the Golden mountain is real ("in some sense") so that, paradoxically, we can deny its very reality and simultaneously reap the effects of its existence in our ability

to refer to it.

The situation is this:

(a) We want a generative theory of content. If Fodor's counterfactuals are relevant to this project, we need to know what positive facts make them true. We need to know this to even see if they *actually hold.* It is intuitive, and fits our semantic expectations, that an asymmetrical dependence holds between buffalo caused tokenings of **cow** and cow caused ones, but who can really say in the absence of an explanation of the facts that underlie the required counterfactuals? The point, put roughly but intuitively, is that counterfactuals by their nature aren't "out there", busy at work in the world. But the physical facts that generate contents, including erroneous ones, are "out there", and all too busy. *These are what we need to get at.*

(b) If the counterfactuals *as representational statements* are said to do the generative work, then we have two problems: The first is the simple fact that misrepresentation can occur even if no mind is or ever has entertained counterfactual representations like Fodor's. Second, if counterfactuals are representations, they are short hand that represents a great many positive facts. Being representations, the activation of a counterfactual representation in the presence of these various positive facts is itself vulnerable to disjunction. Counterfactual

representation X may be activated by set "a" and later, set "b", yielding the content, **set a of positive facts or set b**. But then Fodor's account becomes circular, because he is assumes the existence of *nondisjunctive* counterfactual representations to solve the disjunction problem. But these too must first be extracted from the disjunctive quagmire, for them to be used to solve the disjunction problem for more primitive representations like **Lo, a cow**.

Perhaps Fodor can break out of the circle by applying asymmetrical dependency counterfactuals at the level of counterfactuals used to de-disjunction our representation of cow. But a deeper problem only appears here. What we are looking at is this scenario: We need to cure Fodor's counterfactuals of disjunctivitus at the **cow** level. Call this level C. This requires, if we rely on asymmetry, another level of counterfactuals, C2, to fix C's level. But this will then, if we rely on asymmetry, call for a further level, C3, and so on to infinity, (Cn). At each level the disjunctivitus returns. If so, then we have not solved the problem even at the level of **cow**. If C3 is disjunctive then it cannot account for a nondisjunctive content for C2. But then C2 can't do the work for C. Each level relies on the next higher level to remove then disjunction problem, so if at any level the problem is not removed, it is transferred back to

the lower level, and then to the next lower level, until we reach bottom--that is, until we reach **cow or buffalo** again.

So counterfactuals as statements don't do the explanatory work. Nor are there *counterfactual facts* loose in the world doing the work of explaining how the fission came about--how it was generated or produced. In the generative sense of explanation, counterfactuals don't provide explanations. Patterns of real cause and effect and other positive physical relations do that, and these are what a real generative explanation would like to get at. Fodor's counterfactuals are no more an explanation of perceptual misrepresentation than the counterfactual above is an explanation of nuclear fission. *They are the result of a satisfactory solution, not a solution itself.*

Is Fodor assuming a solution to the choice problem?

We must also ask if somewhere in the background Fodor is assuming the dictum *one token, one content.* I worry that he is, and is *assuming.* His counterfactuals don't provide anything like this general conclusion about representation. Their use in providing a solution to misrepresentation relies on it. Fodor's approach requires strictly constrained contents be attached to representations--such as having, at best, only those contents

that causal interaction of a very restricted kind produces and collapsing these into a single content, or a single disjunctive content. Here's another way of interpreting the contents generated by indication,

> The all in one theory: R represents lots of entities, generally whatever R *has* indicated and *is* indicating ("indicating" in the sense of causal covariance), in all the possible permutations of this. Reality is a representational free for all.

The fact that there are asymmetrical dependencies loose in the world does nothing in itself to show this theory is false. Only in the context of our preferences--preferences like the existence of nondisjunctive content, and the existence of singular content--are asymmetrical dependencies interesting. But what of these preferences? Why put our confidence in them? They are in fact assumptions in need of examination. Consider: In the cow and buffalo scenario this theory generates the simultaneous contents of (1) **Cow**, (2) **Buffalo** and (3) **Cow or Buffalo** (as well as, I imagine, all of the conjunctive combinations of these and then of these combinations, and so on....). This reintroduces a brand new choice problem and attendant problem of misrepresentation. When looking at a Buffalo we incorrectly token **Cow**, correctly token **Buffalo** and correctly token **Cow**

or Buffalo. Do we simultaneously represent and misrepresent? We are blinded to this kind of issue by the unexamined assumption,*one token, one content.* But this is a vulnerable assumption. This goal of the exclusion of all but one content--the goal of solving the choice problem--is never defended and ultimately appears to be misguided. In the next chapter we will explore at length how multiple contents can actually be useful agents in the explanation of error.

2. Teleological theories: Evolved functions and mental contents

The idea behind teleological theories is simple. What a representational state is about has to do with the evolutionary history of the cognitive mechanism that produced the state.[81] Evolutionary history determines biological function, biological function (by reflecting this history) determines representational content. *The hope is that in evolutionary history we have a solution to the choice problem.*

An illustration: Polarization measurements by bees represent sun direction on cloudy days, because allowing these measures is what sensitivity to polarized light evolved in bees to do. This in turn lets them navigate back to flower patches, the key food source, regardless of the sun's visibility. This is the

[81] See Godfrey-Smith's "Misrepresentation" (*Canadian Journal of Philosophy, 19,* pp. 533-550) for further explanation of this distinction.

purpose, the "teleology" of the representational mechanism. Not "purpose" in the sense of conscious intention of course, but only by analogy. Conscious goals liken to states that various operations and features of a system conspire to produce--in this case, "evolved to do". As with conscious goals, these are not always the end states actually reached, but these states have a special status, it's these that are "aimed at".

Cashing out the notion of "evolved to do" can be complex, but the core idea is that ancestor organisms hit upon a genetically underwritten structure that produced certain effects and increased the odds that they would survive and produce descendants with the same genetically underwritten structure, and this explains why present day organisms have this structure and enjoy it's effects.

How can we extend this idea to mental content? The simplest way is this slogan, "What a representation evolved to do is what it means." As Papineau puts it,

> The disposition to form a given type of belief [a representational content-type] is explained by the fact that that belief has typically arisen in certain circumstances and in those circumstances the actions that it has directed have been selectively advantageous. The typical circumstances in question are the belief's truth-conditions.[82]

[82] David Papineau, "Representation and Explanation."*Philosophy of Science* 5, p 557.

This is appealing. How does it apply to misrepresentation? Recall our bees. Their sensitivity to polarized light is part of a number of perceptual/cognitive abilities that produce a representation of flower patch position. We can manipulate their representation by rotating light polarization. We put them into an auditorium with a polarized light, hive in the center, patch in the West. Out come the bees, they find the patch, and return with food. While they're in the hive we rotate the polarized light 180°. This time the bees fly to the East corner where, say, a priest terrified of bees has been cruelly tethered. The bees show up and the person starts screaming for divine intervention. Now we have the choice problem: What is the content of the bees positional representation, **flower position** or **screaming theist position**? The teleological answer is ***flower position***. Perhaps regrettably, sensitivity to polarized light evolved in bees to track down flowers, not theists. To cast things into a disjunction, the content of the bees' navigational representation isn't **position of flowers or position of shrieking persons**.

The bees misrepresented the direction to the flowers. Indeed, you could even say we *tricked* them. The purpose of polarized light sensitivity in foraging bees in is the determination

of flower position relative to the hive. This is what reliance on the orientation of polarized light has evolved to do in bees. Change the polarization, change the direction the bees represent as *the direction of the flowers*. When a properly functioning representational system doesn't perform its representational purpose, it misrepresents.

Limits of the teleological approach

Sterelny points out that the appeal to teleology can be modest or ambitious. An ambitious account would try to ground all representational content in purpose. Papineau's view is an example. The problem with this ambition is simple: What reason is there to think more than a few representational states have anything like an evolutionary history? Not only may the evolutionary history be convoluted, it may be nonexistent. Sterelny writes,

> Most of my beliefs, no doubt, are had for the very first time in human history by me. This is no tribute to my extraordinary and original genius....A large chunk of my beliefs are first person beliefs: Beliefs that I have done such and such. No one else has such an interest in me. [83]

A similar point involves the distinction between

[83] Kim Sterelny, *The Representational Theory of Mind: An introduction.* (Cambridge: Basil Blackwell, 1990) p. 129.

mechanisms and states of a mechanism. While mechanisms--neural configurations or architectural tendencies--may be subjects of evolutionary selection, the states of these mechanisms can hardly be. States don't have a history of genetic selection. But representational content is often attributed to a state, not a mechanism, as the phenomenon of misrepresentations like, "The Earth is flat." shows. No such thing as a "Flat Earth detector" exists as an *evolutionarily-selected* mechanism. Beliefs like the ones Sterelny mentions are surely nonevolved (non-genotypic) states of cognitive mechanisms.

This alone doesn't mean that content can *never* be generated by the teleology of a mechanism and then applied to its states. *Oceanus* crickets have a single neuron (a "command neuron") that when activated in the context of flight causes the cricket to spiral wildly. High pitched sounds--especially those of bats, which prey on flying crickets--activate the neuron. The cricket spirals and doesn't get eaten. So it survives to reproduce similar crickets. Apparently a primary evolutionary factor in the existence and effects of this command neuron (a mechanism) is it's role in bat avoidance. Teleologically speaking, it's activation (a state) while in flight would carry the content, "Bat", at the very least. Not because the state evolved (it didn't), but because the capacity of the mechanism to produce such

specific, selectively advantageous states, did evolve under pressure of bat attacks. The states are not selected for, but the mechanism to produce specific states like the activation of the command neuron are. The world-representing content given by the mechanism's history exists in just these states.

Sterelny endorses only a limited reliance on teleology. On his view basic *perceptual* contents can be fixed this way.[84] Like the role of polarization in the life of foraging bees. This is a perceptual content. Edge detection in our visual system as representing an object's *edge* is another example. A large part of our representational biography will have to be produced in another way. But even this modest use of natural selection has been attacked.

Natural selection and perceptual content

The classic attack on teleological content is found in Fodor.[85] Before we outline Millikan's elaboration of teleology let's look at his challenge. Fodor argues that appeals to biological function don't give us a way out of the disjunction problem, they only introduce a new disjunction problem. His famous black dot detector and "fleebee" example illustrates the

[84] Ibid., p. 138
[85] See chapter 1 in Fodor's, *A Theory of Content and Other Essays* (Cambridge: MIT Press, 1990).

idea: Frogs snap at small, dark, flying objects such as flies. But they also snap at black BBs shot past them with as much zeal. There is nothing in evolution or the basic facts of natural selection that shows us the content of their tongue flick response is **fly** instead **BB**. Frogs have evolved to respond to either just as well. So why not say frogs have a fleebee (**Fly or BB**) detector? Further more, from the frogs' point of view, at best its tongue flick response reflects seeing a Black Dot moving by. The frog has a black dot detector as much as a fly detector. Evolution doesn't determine which is the correct content ascription. If there is any fact of the matter here that evolution can ground, it is only that what has mattered to survival of a frog is how many flies get caught, not what the capture-triggering representation's *content* is. And on this point **fly**, **fleebee** and **Black dot** are all on a par. They all work as well at catching flies.

Fodor's basic point is that what natural selection produces isn't particular contents like **fly or fleebee** but more general results--like motor responses that lead to survival. As he puts it, "Darwin cares about how many flies you eat but not what description you eat them under."[86] So the choice problem isn't solved by evolutionary history. It isn't the sort of thing evolution

[86]Ibid., p. 21

has any bearing on.

There are some obvious replies the friend of teleology could make. As for calling a frog's tongue flick response a Black dot detector, this looks be an *incomplete* description.[87] Black dots are visual experiences and the frogs motor cortex does respond to these. But in doing so, it is detecting what these represent, and these can be flies. If we take evolution at face value, it appears that flies are what the tongue flick response evolved to catch. So it is perfectly consistent that the flick response *also be a fly detector.* It needn't just be a Black Dot detector. Fodor hasn't shown the right content isn't ultimately **fly**.

Fodor's worry that the fly detector is, from the view of evolution, a fleebee detector is a different issue. Sterelny and others have replied that there is nothing *arbitrary* in claiming natural selection has produced a fly (insect) detector in frogs and not a fleebee detector.[88] *Actual* history is the generator of content. The basic environmental factor that underlies the capture-response in frogs is flying insects. It nonarbitrarily excludes BBs as the perceptual content. If the environment had been BB infested during the evolution of frogs, it is likely that

---

[88] Kim Sterelny, *The Representational Theory of Mind: An Introduction* (Cambridge: Basil Blackwell, 1990) pp. 126-7.

further perceptual discrimination would have been deployed to avoid the consumption of large quantities of nondigestible metal pellets. Had flies' appearance changed so that they became less and less BB like, frogs would have changed with the flies, becoming less and less likely to snap at BBs. It's flies, not flee bees, that frog evolution was and is *tracking*. So this is the content of their capture-triggering representations.

To help focus this point, Sterelny invokes Sober's distinction between *selection for* and *selection of*. If we have a screen that will let only objects 1mm in diameter pass through, these all might, as it happens, also be green objects. But the filter selects *for* 1mm, even though it performs a coextensive selection *of* green objects. So it is with the frog's fly detector. The detector was selected for flies, though it performs the selection of coextensive fleebees just as well.

Millikan and teleological content

Depending on the sort of reply Sterelny gives, Millikan exploits evolutionary history in a much more sophisticated way than Papineau. Millikan's approach deserves special praise for it's recognition of the complex issues involved in any biological treatment of representation. It is simply the finest attempt to illuminate content with traditional evolutionary biology that I

have ever encountered. Unfortunately space only allows us to touch upon the basics in what follows. First some technical terminology. Millikan relies on these special concepts,[89]

basic factor

A basic factor is a factor that produces a selective advantage for certain behavior

Normal Case

A Normal Case is a situation where behavior does lead to the occurrence of the basic factors. The Normal Case needn't be the statistically likely one.

Proper Function

Proper Function occurs where the behavior produced does, in fact, lead to the factors occurring in a Normal Case.

Representation Interpreter

A Representation Interpreter is a mechanism that responds to the tokening representation, producing behavior appropriately shaped by the representation.

---

[89] Here I follow Cummins summary in *Meaning and Mental Representation* (Cambridge: MIT Press, 1989), pp. 76-80.

Millikan, like Papineau, sees a tight link, perhaps identity between truth conditions and meaning. Following Cummins, a minimal summary of her theory of content would be,

> T is a truth condition for a representation R if and only if T's occurring is a basic factor within a Normal Case for the performance of the Proper Function of representation interpreters.[90]

Examples are critical to understanding this. Think about polarization's role in bee navigation. Polarization is switched and the bees go in the opposite direction, away from the flowers. Flowers are "what they are after", in the Normal Case. It's the presence of flowers at a certain location that makes the bees' representations of location correspond to the facts: No flowers and the truth conditions for these representations aren't met. Flowers (as *food*) are the basic factors. Recall our recipe for truth conditions; they are the basic factors. If the representation is tokened when some or all of the basic factors are lacking then it is false. It misrepresents because it occurs when it's *truth conditions are not satisfied.* That is, the flowers (and the food they provide) were not where they were supposed to be.

Millikan's subtlety is found in her careful discrimination

---

[90]Ibid., p. 76.

between basic factors, Normal Case and how these support Proper Functioning. So there is one simple issue to face: Granting these notions pick out facts about organisms, can they really support specific content, solve the choice problem and allow for misrepresentation? I'll argue that evolution as a physical process radically under-determines content, so Millikan's approach to error must fail.

Many factored history and the problem of content assignment

Sterelny's reply to Fodor, good as it is, faces a number of problems none too different from Fodor's original challenge. These are problems for any appeal (clumsy or subtle) to evolution in order to fix contents, solve the choice problem and provide for misrepresentation. They challenge Millikan's view in a fundamental way.

Sterelny answers Fodor by insisting on content assignment that conforms to *actual* history, specified in enough detail that *flies* (or at least *edible food particles*) are recognized as the source and so substance of the content. Millikan would agree. Notice that the application of her special concepts to any real world case is determined by the actual evolutionary history of the organism. But if the specific facts of selective history are what matters we have to wonder, "What of the many,

convoluted evolutionary facts? How do these combine to fix content? In what ways?" We still have a long way to go before we reach a clear account.

For instance, while it's true that fly capture was selectively advantageous in the development of frog's tongue flicking by providing *food*, other powerful factors surely exist in the selective history of this response. It is only a question of how far we are willing to go back in the evolutionary history that ultimately leads up to this response. Let's look at some cases that illustrate the *form* of this problem, and then diagnosis the fundamental problem.

Suppose--and this is, for all we *know*, as possible as not--that the original tongue-flick response was to the sudden appearance of other animals. Proto-frogs used the tongue flick response to disorient attackers. So why don't we have a disjunctive content, "**attacker or fly** (or **food particle**)"? Unlike Fodor's example this involves, suppose, real selective history. For an example better grounded in current theory, take the case of vision. Vision evolved from patches of photo-sensitive tissue. It is often thought that these patches allowed the detection of predators above the organism, via the shadows the predator cast. So a drop in light-activation has the content, "**predator overhead**". But another obvious content is

nighttime or **safe dark place**. Each of these have powerful selective advantages. So are we facing a disjunctive content, **predator overhead or nighttime or safe dark place...**?There's no doubt that this kind of complication is real enough. It's no comfort to evolutionary content theorists that putting forward clear, fully established cases of multiple selective factors is sometimes difficult. Quite the opposite. That just shows we don't know what the content of a system's representational states are.

Let's set aside the pressing epistemological problem for teleological theories, "What was the actual selective history of X?" True, we would like theory to give us a fair shot at determining what the content of our representations are, but failure here doesn't refute the truth of a view. Instead my point is that evolutionary history is many-factored and convoluted and this has to affect any would-be content assignment based on real evolution. Is there principled way to wall off many influences, and single out one, and so arrive at a single content like **fly** (or **food particle**)? Given the history and resultant structure of a biological representational system, can this be nonarbitrarily parsed from many, many other factors that helped select for present day frog behavior? Or is the content disjunctive or even blended in some way--67% food particle,

23% attacker, 18% drinking from water in deep cracks, etc...?

One might reply that in order to block out selective factors from the more distant past, we should only attend to those factors within the time span of a particular *species* existence. There are at least three problems with this move. First, in many cases the process of recruiting precursor structures for new uses will occur within the span of a species' existence. Second, we needs some justification for the idea that evolutionary content must be species-bound. This appears to have no independent justification in the nature of representation. Indeed, it looks like we would loose access to a great deal of fundamental contents if we restricted our generating factors to only those occurring in the span of one species' existence. Finally, the notion of a "species" doesn't define an exact boundary in time. Instead, the term is short hand for a region on a branching evolutionary tree of descent and diversification. There is really no nonarbitrary line that delineates a species.

One might be tempted to focus on the recent selective influences and dismiss those more removed in time. The *latest* history supersedes the rest. But the basic question only reappears. Since its effects on genetics and behavior are fairly gradual, at what point does it gain control over the content assignment? Or does it only yield a disjunction or blend

between itself and previous factors? This is simply our original, unanswered question. No progress is made. Other worries: What if the latest major influence pales in survival and reproductive impact (as it most always does) when compared to the previous ones? Why should it receive any special semantic significance? What if the most recent influence is ultimately nonadaptive, while contents from previous ones are adaptive? What if the latest influence is itself a combination of many factors? Within the evolutionary framework these questions appear to have only the most *tenuously* justified answers. But rather than launch a lengthy discussion of these, let's just focus on a simpler and more basic question: When does control of the content switch from the previous influences to the more recent? When has a change in content occurred?

An illustration of the issue: Suppose frogs are placed in a lab environment saturated with a reproduction inhibiting chemical. Flying gelatin pellets that contain an antidote to the inhibitor are shot through the air and tracked and eaten in great quantity.[91] Because of the capture of these pellets the frogs (with their tongue flick response) replicate (for as many thousand generations as you like) in this new environment. Add, if you wish, that differences between the flight properties and

[91] Further suppose that our frogs' nutritional requirements are dealt with via periodic injections.

appearance of flies versus gelatin pellets lead, via selection, to some (perhaps minor, perhaps major) behavioral and structural changes in the frog. Does any of this mean the content of the representation that sets off the tongue flick response has became **reproduction inhibitor antidote**? If in this case we say it is still **fly** because flies, not antidote pellets, originally drove the evolution of the response, then it will become **attacker** in the attacker disorientation case above; frogs today are perpetually mistaken. Alternatively, suppose the tongue structure is radically changed and the behavioral response of the population is dramatically fine-tuned to the flying antidote pellets. This will happen a few DNA base pairs at a time. When is this enough to change the content? At what point in the progressive alteration does the content change? If we let a few flies loose in the environment late in the story and the frogs zap these, when do the frogs start making mistakes?

Consider the resources with which evolutionary theory has to respond to these questions: (1) Actual environmental effects on a population's varying DNA via selection. (2) Actual chemical dynamics of DNA alteration via selection, random walk, chromosome crossing over, gamete recombination and random mutation. There is nothing else. The former only poses these questions and the latter is only the detail of how the

former operates on the structure of chromosomes. These details are profoundly interesting, but they could easily be different and evolution have the same effect on neural structures and animal behavior. So they don't look to be relevant to the fine grained fixation of content. Natural selection alone is supposed to do this. And *that* is what poses our questions. So from the evolutionary perspective no nonarbitrary answer appears to be available. Shall we simply *stipulate* our answers? Then why appeal to evolutionary history at all?

Let's reflect on the moral of these more or less fanciful examples. Assuming we are trying to produce a single nondisjunctive, coherent content, all this suggests a simple dilemma,

(1) Either behavioral/structural alteration via evolution is sufficient to alter the content or it isn't. If it is, there is no nonarbitrary point at which to say the content has switched, since such change is liable to be to some extent gradual and continuous (even granting punctuated equilibrium models of genotypic change). If change isn't sufficient then either we abandon hypothesis of evolution-based content or we embrace an imposed, arbitrary view of the content switching at some particular time in the new environment--we merely stipulate that

after twenty two generations (etc.) the content switches to antidote pellet.

Or (2), we accept a view that blends the content, via disjunctions or some many-entry listing. So it becomes wildly disjunctive (the content is fly (food) or antidote pellet or attacker or....) or ridiculously muddled; a great many different contents at once, each depending on different factors in the evolutionary history, yielding a unified, single, instantaneous whole that looks like the very model of incoherence--p and q and r.... simultaneously, where p entails not q, not r, not.....Or perhaps we face a more refined version of this; a fractional content, but one totally alien to the very notion of meaning. R's content is simultaneously 23% this, 31% that, 19% something else....one that realigns it's fractions in precise steps with the changing of DNA base pairs.

This would be a content that represents nothing in the world, or even in the imagination. Obviously nothing in the world: A horse is a horse of course, not a blend of moonlit shades and a bush and Volkswagon-in-the-dark and etc.. Nor are we imagining something like unicorns, a complex content built out of singular, determinate component contents like **horse's body** with a **long horn**....instead the idea is

something that infects the creature-identity of every part, every division within a fractioned whole. Each part is as much a fractioned mess as the whole, for the same reason--convoluted evolutionary history of the content bearer.

None of these results is palatable.

Conclusion; multiple evolutionary factors, multiple contents

There is *general point* being pressed by these more or less fanciful examples. Most biological mechanisms evolved from precursors that had a prior and very different history of selective factors. It is also common that multiple selective factors worked simultaneously. How could one not expect that to some extent foraging at night among hot desert plant eaters was selected for by *both* safety from predation and the advantages cooler temperature offers, including water conservation, less demanding internal temperature management, the increased water consumption gained through ingesting dew and so on. Either way, once we take seriously the claim that *actual evolution* determines content, there will always come a point in the evolutionary history of any organism or organic structure where evolutionary content of representations (like those that respond to darkness) becomes divided between distinct selective factors, introducing multiple

contents, disjunction or incoherence.

I believe the conclusion to be drawn is simple: All theories that draw on evolutionary history to fix determinate content fall victim to this basic dilemma: Either the single content is not particular--it means many things at once, either absolutely or in some fractioned whole--or it is singular but radically disjunctive. Applied to Millikan's view, the point is simply that evolution doesn't array basic factors, Normal cases and Proper functions in any way that escapes this dilemma. There are too many, or alternatively, too convoluted "basic factors" to pin down anything palatable as *the* content.The choice problem looms everywhere and remains unsolved.

Dretske and teleology

Dretske, a pioneer in the subject of naturalizing representation, is famous for his early work on indication theory. In response to the problem of misrepresentation and the failure of his earlier attempt to account for error via a "learning period", he introduced a indication/teleological hybrid. I will only sketch to the point where it is apparent that evolutionary factors are necessary for his view to succeed, and then reapply a version the critique of evolutionary content-fixing above. I apologize for the repetition this involves, but Dretske has been such a

seminal figure in recent discussions of representation that I believe he deserves to be addressed separately.

Dretske offers this account of content as *functional meaning*, where function is determined by evolutionary history,

> Functional meaning (Mf): d's being G means$_f$ that w is F= d's function is to indicate the condition of w, and the way it performs this function is, in part, by indicating that w is F by d's being G.[92]

Dretske's own example is wonderfully chosen. Certain bacteria are sensitive to magnetic fields. In the Northern hemisphere this sensitivity allows them to avoid lethal O_2 rich waters by swimming along the northern direction of the field lines. That sends them down, toward O_2 deplete waters near the bottom of the ocean. Put the same bacteria in the Southern hemisphere and they will travel northward along the field lines, killing themselves by swimming up into O_2 rich surface waters.[93] Back to our definition of Mf: What does the magnetic sensitive response (d's being G) indicate? Indication as merely reliable correlation yields a plethora of answers: The way north, low O_2 waters, the alignment of magnetic field lines, the direction towards the nearest iceberg, the pole-star, the oil tankers daily going from the North sea to a refinery in Newfoundland, etc..

[92] Fred Dretske, "Misrepresentation." In R. Bogdan (ed.), *Belief: Form, Content and Function* Oxford: Claredon Press, 1986, p. 33.
[93] Ibid., p. 26.

The most reliable indication of all is simply the *disjunction* of all these. So pure indication is infallible, leaving no approach to the choice problem or misrepresentation.

That's why Dretske includes a functional (read: teleological) element. The indication powers of d are narrowed down to one relevant content, the one which it is d's evolutionary function to indicate. In the bacteria's case, low O_2 water. Why? Because this is a major selective advantage of magnetic sensitivity in these bacteria. So his view is just the evolutionary solution to the choice problem already discussed and critiqued.

To deploy our critique against the bacteria example: Put the bacteria in a lab, alternate the magnetic field and start selecting for responses that are different than those in the wild. How about motion *perpendicular* to the field? We keep the water uniformly low in O_2 but add new advantages to changing the response to perpendicular swimming, like higher food concentrations in that direction. Some bacteria--mutant types--that naturally deflect from field-parallel swimming reproduce more than others. The more they deflect, the more they reproduce, until new genes start to dominate the population. As this change gradually proceeds, all the questions we asked about frogs and food pellets reappear, along side any questions

about the larger history of the magnetic response. It is credible that there were previous epochs in natural history when the response, or precursor of the response, was selected for by factors other than O_2 concentrations. It's just as possible that these factors are still at work today. Dretske's view is as vulnerable as Millikan's, for the same reasons. Indeed, the dilemma outlined above is equally potent against any evolutionary approach to solving the choice problem.

Evolutionary content and process/relationship relativity

The problem with teleological content explored here is based on the goal of generating one content for each token. This is impossible, and so teleology fails to solve the choice problem. Because it fails to solve the choice problem it fails to solve the problem of misrepresentation--it doesn't "get off the ground". But it is important to notice that on process/relationship relativity, evolution (as a species of causation) can be called upon to generate coherent contents like **fly** or **flowers**. If there is a causal/selective process that maps the existence of flowers in the distant past with the existence of various neural structures (via genetics), we can identify a content of a state of activation of this neural structure relative to this process, **flower**. This will be explained in fuller detail in chapter V. For

now note none of this involves any special appeal to *purpose*. It is only an expression of causal content of a particular complicated kind.

3. Clark's "noise category" response to misrepresentation

We've just finished a critique of the standard contemporary approaches to content and misrepresentation. But our review wouldn't be complete without a look at a novel and compelling approach recently offered by Austen Clark.[94]

Clark's view rests on classic information theory and the idea that the preferred interpretation of a representation's content is that which maximizes "information" available to the cognitive system. You mistake a leaf blowing in the wind for a mouse. Is the content of the activated representation **mouse** or **leaf or mouse**? Again, we face the choice problem. Clark's basic solution: If we aim for the greatest quantity of information, we will treat some inputs as *noise*, leaving representation's content nondisjunctive. Occasional leaf-caused activations of the mouse representation are noise. This leaves the singular, nondisjunctive content of the representation intact, and activation of this representation by nonmice count as

[94] Austen Clark, "Mice, Shrews and Misrepresentation." *Journal of Philosophy*, (1993), pp. 290-310.

misrepresentations.

What justifies the assignment of leaf-caused tokening to the "noise" bin? Compare the two cognitive systems: One system treats these inputs as "noise", the other fails to introduce this distinction. Given simple axioms of information theory, it is easy to show that the system that is able to treat some of the representation-activations as noise can reflect *more information* about it's environment than the other system. If you have a wholly natural system that is able to make the noise/veridical distinction you have a system that solves the disjunction problem, because a basic aim of cognitive systems is to increase the amount of information they gather from the world.

But how does a system decide that some activations of the representation count as "noise"? Clark's answer couldn't be more intuitive: The system takes a second look and discovers that the would-be mouse isn't a mouse. It's a shadow, or a small brown leaf trembling in the wind, or whatever. Closer, repeated examination drives the deployment of the "noise" category, and the information the activation of the representation carries increases dramatically. No longer ambiguous between a wide range of environmentally present types, the representation now occupies a role in a system that is highly indicative of one type,

mouse. This steady increase in information carried by the representation's activation is what *justifies* us in saying that the content of the representation is not disjunctive, but specific to a single type. When non-mice activate the representation, the system has made a mistake.

This is a very nice account. It fits the learning and adaptation we expect of successful organisms without invoking any exaggerated ideas about evolutionary history directly fixing nondisjunctive and coherent univocal contents. It draws upon a fruitful and mathematically rigorous treatment of content-states. It's the *looks-like-science* sort of thing that isn't hard to imagine wholly physical systems implementing.

But while granting all this, I don't believe Clark's view is an account of the *generation* of misrepresentation. The core shortcomings are (1) the vulnerability of the noise treatment to the disjunction problem and (2) a complicated version of saying a representation R has content **C** because we intend it to. This would be a circular account, because intentions are also contentful.

The choice problem, ala disjunctive contents, returns

Critical to the account is the concept of noise. Certain representation activations count as noise if reexamination of

the distal source fails to produce R's activation again: We think **mouse**, look again and think **leaf**. The event of a leaf's activating R gets trashed into the noise designation. But how are we to understand the action of making an activation of R become "noise"? It looks like we are facing a *categorization* of this event. In this cognitive context the deployment of a "noise" category itself involves representational content, because on information theory the deployments of categories are representational when they carry information about (indicate) an event. In this case the information is that event x was a noise event, the informational content being **noise**. So when our mouse representation R is activated by a leaf, R is also indicating noise. Here the disjunction problem simply re-arises. We ask, "Why then is the content of previous, veridical R activations not **mouse or noise** instead of just **mouse**?" The more indicative of mice R becomes, the more indicative of noise, since the two increase proportionately.

Perhaps this disjunction doesn't carry as much information as **mouse** alone, but there is nothing in the actual event of reexamination and subsequent deployment of the "noise" category which determines that the content is only **mouse** and not **mouse or noise**. Moreover, a reexamination of a distal source isn't able in the leaf-activation scenario to

decide whether this content is **mouse** or **mouse or noise**. We will only discover a mouse or a leaf. Neither gets us out of the **mouse or noise** disjunct.

Clark may attempt to reply to this concern by insisting that the "noise" category is representationally *blank*--it has no content associated with it whatsoever. But the problem is that "noise" is a category, one that indicates and so on Clark's preferred view of content-generation, has a representational content. As such, in this context it represents at least this: Distal causes not signaling mice. This is a state of affairs, a fact about the signals it receives. Such facts are represented by the category **noise**.

More generally, how does the system ever get itself into a position where it can toss some contents into the "noise" category but not others? Imagine the initial scenario: We are looking for things that are in fact mobile small mammals, in English, "mice". But let's not grant all that baroque content to the system from the get go, let's be minimal and pretend we're a crawling baby. We have had a perceptual episode caused by a mouse, and assign a prelinguistic type to the memory just produced, **M**. We initiate goal directed searching for **M**, and we have a second perceptual episode, which we categorize as type **M** also. But focusing on the same spot reveals a different sort of

pattern of perceptual input, one markedly different than **M**, a flat, relatively immobile elliptical object. We type this differently than **M**, we type it as **L**. Only at this point do we have the minimal scenario for then deploying a noise category. We can inspect the leaf and detect error.

But a problem immediately appears: If the original content of R is caused by some mouse scurrying by, we get to start with **M** alone as R's content. But this doesn't help, since the first leaf to causally activate R will turn R's content to **M or L**, *prior* to the reinspection of the leaf. Deploying a noise category first requires that we discover some of our perceptual episodes are *in error*. And doing that first requires we have a nondisjunctive content, **M**. But before we can use the noise category we already have disjunctivitus. *It looks like we never get into a position to use the noise category.* So we never get to deploy its information increasing powers.

Projecting our theoretical aims onto a system

The claim about preferring those treatments of a system's representational states that maximize information content is also troublesome. A principle of charity in our evaluation of cognitive engineering is being offered. Perhaps well adapted systems typically benefit from an increase in the information

content of their representations. But are they in fact as generously endowed with this increase as we might hope? The fact of the matter seems independent of how we treat a system, especially of *how desirous we are that the system be one that can make representational mistakes.*

Is there an instrumentalism about cognitive ascriptions being assumed in Clark's account? The main problem with such an instrumentalism is that the observer-independent facts of cognition are replaced by ways observers might describe a system. Cognitive facts are no longer facts existing through the system itself and its situation in the world. They become relational facts between it and possible observers, contingent on the ways observers *choose* to describe the system. The sort of question such views flounder on is simple: Is the system *really* nondisjunctive in any of it's contents, regardless of how we conceive it to be? This is just our explanatory task all over again.

Worse, the idea that an activation episode is *reassigned* to another token (the noise category) seems either to be an unreal abstraction in our description or some sort of magic, because the physical facts remain: R was activated by both **M** and **L**. We are not moving **bolded words** from one bin to another. We are dealing with irreversible history. I do not see

what it literally would be to have the noise category do what it is suppose to. If we simply claim it is, what makes that claim true? If we grant that "charitable" understanding does fix the "facts" here, we still have this problem: What of us? Our charitable treatment of a system as deploying nondisjunctive contents itself relies on our contents being non-disjunctive. We are asserting that, "System S is misrepresenting" not, "System S is disjunctively representing". If we explain our ability to treat a system a particularly way by invoking the concept of "treating" in reference to the very act of treating a system a particular way, then this act of treatment is itself subject to the same issue--is it nondisjunctive, with a content like, **treating S as misrepresenting or treating S as disjunctively representing**? We're off to a regress. "Charitable understanding" is as contentful, and as vulnerable to the disjunction problem, as any other representation.

Further, what makes the noise category be about noise? When a system "treats an activation episode as noise", is this anything intrinsic to the system? But what could that be? As far as indication goes, a great number of things activate the noise category, other representations and via them, the things in the world. Why isn't the noise category just a second level representation that is wildly disjunctive? And why believe that

any of its disjuncts include the one called "noise"? Is it merely our preferred description at work here? But then that is not a fact about the system, but an outside stipulation.

One also wonders just what exactly noise *is*. What kind of entity is this? In what sense is it admissible in a naturalist ontology? If we simply claim it is "those activations of a system's representations that reduce information content" then a lot of things that are not noise will be counted as such--like disjunctions used in logical inferences. If our view is that noise is any information-reducing activations which we *desire* to exclude, we are assuming contentful states (desires) at the start of our explanatory enterprise---a circular view.

Whatever we make of the nature of noise, there remains the question of why we should consistently view cognitive systems as *information maximizers.* What facts in the world underlie this estimation? Clark claims that systems evolve to be information maximizers because of the selective advantage. We have already noticed that it is hard to see how this advantage ever gets started. The adaptionist picture is also murky, because in many scenarios too much information impedes action and only simplifying assumptions allow us to proceed. Often information *reduction* is the superior value: Imagine a shrew and mouse sighting cult: We wander the fields,

and see lots of both and feverishly distinguish them. Being vegetarians and having no truck with a university job as a biologist, we gain no real advantage in the world by deploying a noise category to do our distinguishing; in fact our distinguishing is a dysfunctional obsession--always trying to see if we are right in our initial judgments, to see if that little brown thing *really* is a mouse or a shrew, distracts us from food and sleep. Yet we *do* feverishly distinguish the two, even though it is a waste. What reason has Clark given us to say in *this case* we are best understood as misrepresenting from time to time? It can't be that our information increase is *better for us*. It isn't. More is not always better. This is obvious in physical science and enjoys endless real-life manifestation, too.

Finally, the most important worry about information maximizing takes us back to the original choice problem. An information maximizing interpretation of the content facts doesn't show that unwanted contents (like the disjunction **M or L**) are no longer present. Information maximizing interpretation doesn't have the ability to *screen out* other contents, but only draw our attention to certain contents and not others. Relabeling some contents "noise" certainly doesn't screen out any of them. It is an *ascription* guide, not a screening property or a principle of generation. Like Fodor, Clark seems to be

assuming a solution to the choice problem when he has nothing like it.

B. Conclusion: The Choice Problem and the Failure of Current Naturalist Theories.

All the leading contenders available today in the Naturalist tradition fail to solve the choice problem. *This represents the failure of a tremendous amount of intellectual effort by some extremely talented human beings. Why have they failed?* The answer and their exoneration is simple: No successful solution to the choice problem exists. The choice problem is *not a problem*. Philosophers are fond of pointing to a unifying defect within a period and domain of philosophical effort. Let's indulge. The period in question is the latter Twentieth century, the domain mental content and misrepresentation. The defect I see is one of *conservativeness* and the quest for *unreasonable simplicity*. The perspective of psychological ecology suggests that what we have called "the content" of our thoughts is in fact a gloss for a vastly more complex and subtle embedding in our environment. Both conservativeness and unreasonable simplicity are reflected in an unwillingness to accept the wild, frightening multiplicity of content most any representation

enjoys and work within the implications of this.

All prominent, contemporary theories make the same mistake: Fixating a single, privileged content is their starting place for solving the problem of misrepresentation. The resulting disaster is now familiar: Disjunctive or incoherently blended contents appear when we encounter a system-history where different activation episodes have different source objects (or properties) involved in the same representation. We take these source objects/properties as contributing to content and sum these contributions up via a disjunction or blending. The task then is to isolate one of these source objects/properties as providing *the* content. This is the choice problem; how do we justify choosing one, non-disjunctive content out of all the contributing contents? Each of the theories above tries to provide an answer to this problem. Their methods for choosing one content over the others collectively fail.

A diagnosis, a hypothesis

We need to admit multiple contents for most any representational state, none of which are intrinsically more or less correct. That is, admit that the choice problem is unsolvable. Instead of vainly seeking a solution, we can work with multiple contents and understand misrepresentation.

Process/relationship relativity let's us do just that. I'll illustrate how next.

Chapter V. A Look at Process/Relationship Relativity in Action

To begin our exploration of process/relationship relativity, let's start with a charitable stipulation. Let's simply assume that representational content is relative to physical process/relationship types. We will interpret this in an intuitive way without demanding detailed hows and whys that lie behind the content assignments process relativity makes. These more technical issues will be explored in the following chapter. Process/relationship relativity will simply be an incomplete but promising hypothesis at this point. This chapter only aims to show the general attractiveness of process/relationship relativity in the face of challenging problems, ones that plague all theories of content.

In the following I will argue that process/relationship relativity can easily solve the *disjunction, depth, qua* and *misrepresentation* problems. Process/relationship relativity offers simple paths around these by rejecting the choice problem as a *problem* and replacing this "problem" with a tractable task. We'll begin with a very obvious and reasonable theory of perceptual contents, the "causal theory", and use it to provide illustration.

A. Three Basic Problems for a Theory of Content: Disjunction, Depth, and the Qua Problem

Let's assume that causal relations between the world and ourselves can produce representational content. To the extent that this assumption allows us to explain the generation of content and misrepresentation, it will be justified. There is a lot going for the idea from the outset, too. Ordinary examples like photographs, recordings and video support it. What they *are of* is what reflected the photons onto the film, etc.. At least so it appears in our ordinary way of thinking. They are not merely interpreted as being of these things--indeed, our interpretations can be mistaken, "Oops. I thought that was a picture of a tree, not a celery stick.". The same is true of our perceptual experiences, at least usually. What we perceive is what is causing our sensory organs to activate in particular ways. It's hard to understand our sensory system's representational power in any other way. And this all seems like a natural process. Here's an irresistible opportunity to "naturalize" at least a segment of representational content; *perceptual content.*

But starting with a simple theory of content like this,

> Representation R has content C if C causes R to be tokened in a cognitive system.

the misrepresentation/disjunction problem occurs in the familiar way,

> We have a representation that is activated in the presence of horses. The sight of horses causes us to think **horse**. Suppose at times we also make a mistake--misrepresent--and think **horse** when we see burros. Our representation doesn't have the content **horse or burro**. It has the content **horse** and we've misrepresented burros as horses. But both burros and horses cause **horse** to occur in us. Both of these are types that cause the representation to activate.

So we might think that on a simplistic causal theory, our representations of things like horses become representations of disjunctions as soon as another type activates them. And that's the problem, they don't. "Horse" means **horse**, not **horse or burro**.

It's easy to misunderstand the proper target of this objection. Three possibilities: (1) it's an objection to the view that causal relations between the external world's objects and representations can play the essential content-fixing role in the production of content. That is, being caused by horses doesn't play the pivotal role in fixing the content as **horse.** (2) it's an objection to the view that simple causal relations alone are sufficient to give an adequate generative explanation of features of content like the nondisjunctiveness of our

representation of horses. Instead, a theory of perceptual content needs more. This is Fodor's approach, via his theory of asymmetrical dependence. (3) it's a consequence of a suppressed assumption; that only one content must be associated with R. Since picking either **horse** or **burro** is arbitrary and the content can't be **horse and burro** (no such thing ever activated R, since no such thing exists) the content must be **horse or burro**.

The disjunction problem doesn't *show* (1) because the possibility of reading the problem as (2) or (3) remains.

The Qua ("as") Problem

"Qua" is a Latin term that best translates, "as". The problem is this: A single external object can be taken to fall under several different type-contents.[95] A horse is a horse of course, but it is also a *mammal*, a *biocellular system*, a *short term negative entropy system* and perhaps a *collapsed wave function* of some horribly complicated form. All these types simultaneously activate the representation. But when my grandma looks out at her horses, she doesn't represent horses as any of these other kinds. She has no inkling of them. What is

[95] This problem was first noticed by Michael Devitt and Kim Sterelny in their book, *Language and Reality: An Introduction to the Philosophy of Language*. Cambridge: MIT Press, 1987.

it that makes Grandma represent horses without tacitly indulging in modern biology or quantum physics? What makes her representation a representation of horses *as* horses?

Sterelny and Devitt first introduced this problem. Sterelny claims that a satisfactory solution still eludes naturalism.[96] This isn't surprising, since it is an incarnation of the *choice problem*, and the choice problem is a *problem* for Sterelny and Devitt. But just as with the disjunction problem there is nothing to suggest that causal relations between R and the external world don't play the key role in generating the content--whatever we end up saying it is--of grandma's thoughts. For us the question here is how process relativity can respond.

The Depth Problem

The *depth problem* is first cousin to the *qua problem*, it just moves the problem into the system itself: Light reflects off the surface of a horse, strikes our eyes, propagates through the vitreous fluid and triggers higher energy states in molecules on the surface of our retina. Sufficient numbers of these events activate an array of retina cells that initiate a pattern of neural depolarization in the optic nerve....and eventually an image occurs in the visual cortex. Is that image *of* horses? Or is it a

[96] Kim Sterelny, *The Representational Theory of Mind*. Cambridge: MIT Press, 1991, p. 117-118.

representation of retina activation patterns? Or is it a representation of optic nerve firing patterns? What fixes the depth of our representations in the long causal chain from Silver to our visual cortex?

The similarity to the *qua* problem is also plain when we look to our initial replies. They are essentially the same. We don't have an argument against causal content. We have--given what's been said--either an objection to a simplistic causal theory *or* to the view that our visual cortex processes are only representations of horses, and not intervening events too. That is, we have just another incarnation of the *choice problem.*

With the importance of these objections made clear--they offer the theoretical options, not obvious reasons to abandon the fecund idea of causal content--let's explore the process-relative response.

Process/relationship relativity and the disjunction problem

Because process relativity rejects the choice problem as a problem--it accepts the multiplicity of most any representational token's contents--the disjunction problem doesn't arise. In the context of our simple causal theory, different processes produce different world-representing contents. Zebras in the moonlight set off the same representational token as do horses.

But these are different processes, because the causal components involved in the processes differ in this relevant way: They are different species of "equine mammal" involved, one is a zebra and the other a horse. Because this content is process relative, relative to these different processes we have different contents. If we assume that R has a single content, then we must try to unify the items of its activation history into a single content or choose one only as *the* content. The sanest way to unify is via a disjunction because conjunction gives us a contradiction; C becomes **horse and zebra and cardboard cut-out and**.... But with process-relativity's admission of (indeed, *insistence on*) multiple contents there is no need for unification of R's activation history into a disjunction of everything that ever activated it nor to choose a privileged content. Process-relativity easily side steps the disjunction problem. Given the exhausting and futile struggles against this problem in recent philosophy of mind, this is an excellent reason for representational theorists to take a serious look at process relativity--in one simple move it brushes aside what many theorists took to be a mountain.

Process relativity and the depth problem

As in the case of the disjunction problem, process

relativity escapes the depth problem by distinguishing different processes. Here the processes are more or less extensive depending on how far they take us from the representational token. The distal cause as R's activator is a process involving objects and physical events outside of the cognitive system. The retinal image as activator is a process that omits the distal cause, but is a subprocess of the larger process that includes it. And so on as we move closer and closer to the activated representation. Since these are different processes, process relativity can associate different representational contents with each. So there is no depth *problem* here, because we don't assume R can have only one, distinct content. We argue just the opposite: All these contents and more are had by R, each relative to a different physical process R takes part in.

Process relativity and the qua problem

Process relativity allows essentially the same simple and direct solution to the *qua* problem. But here we only appeal to the *type* of the process, not the token. Different types, taken as objective properties, can be had by the same object. A cat can be a mammal, an endothermic animal, a wave form, a protein particulate and so on. So while the *same object* is involved in all these types activating the same representational token,

different *types* of physical processes can be distinguished, all present in the same *token* of a physical process. The process-type of being activated by a protein particulate is not the same process-type as being activated by a mammal, even though the instance of the process is the same in both cases. The same process token can fall under different types, just as the same object can. So here we attend not only to different process instances and their associated different types, but to different types embedded in one and the same instance. This symmetry isn't surprising, since it is different types embedded in one and the same object that give rise to the problem. So naturally we exploit different process-types they give rise to in one and the same instance of a process. The larger metaphysical lesson is simple: What goes for object-types within one and the same object goes for process types within one and the same process.

The *qua* problem is really very similar to Searle's concern with "aspectual shapes".[97] To use Searle's preferred example, representing x *as* water and not H_2O is to adopt a particular aspect on it, to invoke an "aspectual shape". That is, to represent x *as* a particular kind (water) and not as another coextensive kind (H_2O). So a solution to the *qua* problem should also allow us an approach to differentiating aspectual shapes.

[97] See chapter 1 and the discussion of Searle's views on representational content ("intentionality") and its reliance on consciously experienced aspectual shape.

Again, Searle's view is that aspectual shape pervades representation, and only relative to conscious states can aspectual shape be made out. A sketch of his argument,

1) There can be no content without that content having a particular aspectual shape.
2) So all content has aspectual shape as a necessary condition for its existence.
3) Nothing nonconscious has aspectual shape while nonconscious.
4) Aspectual shapes do exist when conscious.
5) So, all contents exists only when provided with an aspectual shape by their actual (or potential) conscious manifestation.

Process and relationship relativity doesn't see it that way. It is very simple to distinguish the difference between representing X as **water** and as **H_2O**. If there really is a difference in representing here (and there might not be) it requires that **Water** and **H_20** are different types had by the same stuff. If this is denied then, since H_2O does exist, water doesn't exist (!) (or vice versa, as you wish). If there are different types here then the process-type of causally interacting with one can be distinguished from the process-type of causally

interacting with the other (it is possible that one and the same causal process fall under two types ("interacting with water" verses "interacting with H_2O"). Again, this really isn't that odd: One and the same object can fall under two types: "Apple" and "fruit". Content in these situations is relative to different types of physical processes, and relative to these processes, different contents are had. These involve "aspectual shapes" exactly distinguished without mention or assumption of consciousness.[98]

B. Process/Relationship Relativity and Misrepresentation: A Solution

The disjunction, depth and *qua* problems are easily set aside by process relativity. This easy, elegant success makes even more attractive the idea that there is a kind of multiplicity loose in our representational contents, but that these contents are singular and determinate relative to the different process-types they are associated with.

Misrepresentation is another matter. Process relativity let's us accept it, via multiple contents associated with the same representational token, but here the account is more complex

[98] Searle cannot use this approach because of his commitment to Internalism.

because the nature of error is.

To understand the basic case of misrepresentation, perceptual error, we need to understand something of the processes involved in external world perception and what process-relativity as an explanation of perceptual error can offer. I believe the assumptions behind the choice problem are correct in one respect, that error occurs relative to a single content. So what we need is,

Process P has certain properties (say, certain causal links to the external world) and these generate determinate external-world content.

And this would be just the sort of thing a solution to the choice problem would have given us. A theorist armed with a solution to the choice problem would have used it this way: The content of R is only this, but *that* activated R. So we have an error.

But notice that this content needn't be the *only* content of R for this starting place to be available. The process-relative elaboration of R's situation is straight forward: *Error occurs when the content bearer embedded in process P is activated, but the external-world object O causally responsible for the activation (if any) is not the object of the content produced relative to process P(the content* ***c****). This object is involved in a*

*different process type, P', so another content (**c'**) is produced via this different process. Relative to the content generated via process P (**c**), this content (**c'**) is in error.*

Let's review some of the key assumptions behind this abstraction and then we'll look at how it works in actual cases. Again, the account will follow this basic plan: Different processes will produce different contents for the same representational token. One of these contents will be elevated to "true" relative to a process, the other will be lowered to "false" relative to that same process. More precisely, we will try to show how the following critical points can be had in a process/relationship relative framework, and that together they generate *misrepresentation*. They are,

1) A contrast in perceptual processes that both activate the same representational token but endow it with different contents is the starting place for a generative explanation of error. Error is an event that critically relies on a *contrast between contents*.

This may seem surprising, since we typically assume that because a representation has a single content, it is the *misfiring* of this content in certain situations that produces error. Process relativity departs from this assumption by attaching *different*

contents to the correct and mistaken activations of the representation. This is a basic departure from the usual way of conceiving perceptual error. Instead of the "misfire" analogy, misrepresentation is seen as resting on contrasts *in the contents given by a representation's different activating processes.*

2) The error producing contrast in the perceptual processes' contents is more subtle than a mere difference in content because two contents may deviate in this way but not be *incompatible* (incompatible perceptual identifications verses mutually compatible description of the same object, for instance). The contrast must be one of *incompatibility. The contents must not both be true of the same object.* In our example above this played itself out as contents that represented different objects, one O, the other not-O.

This is important because it allows the process relativity approach to distinguish merely different descriptions or ways of representing an entity from correct/ mistaken ways of representing it.

3) The concept of error presupposes a standard of correctness that can be *deviated* from. This deviation involves a *contrast* between deviating and nondeviating representations and there is an *asymmetry* in the contrast; one is the standard, the

"correct" and the other is the "incorrect". Again, the truth of the standard content must imply the falseness of the contrasting content if there is to be an error.

This brings us to the heart of the content-contrast approach being taken here. If error is to be generated by certain contrasts in content, these contrasts must be the sort that render one content mistaken.

The proposal is simple: If we can account for these features we have captured what's essential for something to be a misrepresentation. Can process relativity capture these features? I believe it can. The following cases aim to show this, as well as illustrate the simplicity of the approach.

Case 1, misrepresentation in the absence of a distal cause; misrepresentation in principle

Imagine a cognitive system that has never represented Chairman Mao before, but one fully capable of the deed (much like our own). Next take the process ("A") of light rebounding off Chairman Mao, onto the retina and the retina activating a complex process in the visual cortex; an image of The Chairman . This generates the content **Chairman Mao** via causation. Now imagine that the system undergoes another

process (B): *Freak spontaneous activation of an identical pattern of retina and cortical activity.*

The activation of the rods and cones is the same as if there is a Chairman present, but no Chairman. The visual cortex image is the same. The retinal, optical nerve and visual cortex process is the same. But that's where the similarity stops. The process that includes the Chairman isn't the same type of process as the one that doesn't. At best, process B has generated the associated content, **neural discharge of such and such parameters**. While A is a plausible candidate for having generated the associated content, **The Chairman**.The real Chairman was, via the retinal system's interaction with light rebounding off him and propagating via the intervening space, causally responsible for the visual cortex activity (the bearer of content). In this case representation occurs because the real object itself is part of the process. On our simple causal theory, this is sufficient to make R have the content **The Chairman**.

So there are differences open to identification with two different contents, one a "freak discharge" and the other a perceptual representation of The Chairman. In principle, it is possible to differentiate the two so that one does, and the other doesn't, represent the Chairman.

Process B wasn't the same process type as the

perceptual process with the Chairman as a component (process A). Thus difference in content exists. That's the first feature of misrepresentation. We also have two contents that deviate one from the other in an *incompatible way:* The chairman is an external world object, a person in fact, not a retina or other neural discharge. He cannot be both. That gives us the second feature of misrepresentation.

Now we need an explanation of the correct/incorrect asymmetry. On process relativity this is straight forward. *Relative to the process of representing external world objects of the kind Chairman Mao the content of process B (**neural discharge of such and such parameters**) is false and the content of process A (**The Chairman**) is true.* Content is relative to a process as is the event of error. The asymmetry is relative to one of the contents and hence to the process that generated it.

This may appear puzzling because no account has been given as to why one content is the truth-standard and the other isn't. When I wrongly think a moonlit zebra is a horse why is **zebra** the content relative to which the activation of **horse** is an error? What makes **zebra** *the correct* content? The response is that, as far as the contents themselves go, absolutely nothing does. Again, process/relationship relativity

departs from our assumptions. Indeed, with this question we face yet another incarnation of the *choice problem*. So naturally, the question presupposes an understanding of the problem of misrepresentation not shared by our approach. The problem is to show how error is generated and for process relativity this doesn't require an explanation of which asymmetries we *attend to*. That is, the problem is not to also offer an explanation of how an organism *perceives* errors. That's a different issue. I imagine which process/relationship type we elevate to the status of generating the "correct content" is entirely a matter of which processes/relationships we are "interested in accomplishing". Relative to the process of representing only freak neural discharges the content associated with process A (**The Chairman**) is false and the content associated with B (**Freak**....) is true. The multitude of ecological, cultural, evolutionary and neurological factors a word like "Interest" covers only concern which errors we verbally or otherwise attend to. Not what error *is*. By itself, error is multiple for any pair of incompatible contents of the same representation. Each can be the other's error. It is the asymmetrical contrast relative to one representational process that isolates one error from another.

This doesn't mean our attention to some asymmetrical

contrasts instead of others is unprincipled, frivolous or even a bit *parochial.* Often the asymmetry that matters is one that other processes in an organism rely on. For instance, the *truth* of a perceptual process may be important for such activities as feeding, mating, catching, jumping and so on. After all, why do we say that when a frog swallows a flying BB, she has made a mistake? Simple: Because she was trying to *eat.* In many cases our preferences about which direction the asymmetrical contrast goes are actually expressions of biological processes that our representational episodes are embedded in. This embedding in a larger organism is important, if for nothing else than that it allows us to explain how error is not committed by the contents, the environmental relations they reflect or the brain processes involved, but by the *organism.*

It is true that the contrast view of misrepresentation multiplies the occurrence of misrepresentations. For every content there is almost always another relative to which it is false. This does not render the view, or the existence of misrepresentation, "vacuous" or useless. In fact it leaves everything just as it was. The errors that play a prominent role in our survival and daily existence (as described from a human perspective--washing, eating, driving....) will still be errors and will still be the errors we will attend to. The existence of a vast

cloud of irrelevant occurrences of misrepresentation in no way undercuts the fact that these are cases of misrepresentation or any usefulness misrepresentation has in our lives or theories.

Ascriptions and multiple linguistic contents

It is important to recognize that what is being offered is an account of the *existence of error*, not a theory of what guides or produces our *ascription of error* to a system. These are very different issues. But the fact that our ascriptions are linguistic, and so themselves representational affairs, when combined with a multiple content view produces an interesting issue about how we succeed in *ascribing* error.

The issue is this: On the assumption that language expressions derive their content from prelinguistic representations ("purely mental representation"), how do we account for the fact that we can linguistically identify the content produced by one process as the one relative to which content produced by other processes is in error? The problem is that if content is multiple, then the content determined for our linguistic expressions is also multiple. So when I say, "relative to food consumption processes, the content 'BB' is a misrepresentation" I have created an event (linguistic in kind) that represents a great many different, perhaps even

incompatible, things.

This is really an issue in the philosophy of language, but it is worth commenting on in passing because of the apparent tight link between linguistic representation and more basic, nonderivative, nonlinguistic representation. Once we admit multiple contents, this will surely condition our view of linguistic events and their representational content. To successfully distinguish two process types we need to have the content of our linguistic event represent one process and not the other. This is entirely possible with multiple contents, even *infinite* multiple contents had by each representational token (here a linguistic event), since even infinite sets need share no members. In this way we can represent one process without representing the other in our linguistic behavior. We can distinguish the two and present only one as correct.[99]

Case 2, realistic cases of misrepresentation

The "in principle" case is valuable because it illustrates the basic process-relativity strategy. But "freak activation" cases of error are far from typical. Let's explore process relativity in a

[99] On a view like this we are going to see important changes in how we understand language and its use. Exploring this is a project in its own right. But the essential idea of *content communication* remains intact on a multiple content view.

more realistic setting. Once grasped, process/relationship relativity has a simplicity and intuitiveness that lets us understand the basic nature of our day to day experiences of error and fallibility with ease.

Another goal of this chapter is to relate processes relativity to some rigorous, mathematically well defined models of learning and cognition. Among these are new connectionist models, in particular the *Kohonen feature map* and *computational models in cognition*. This will show how easily and naturally process relativity *fits into* and *interprets* ground breaking work in cognitive science and computational/representational modeling. It's hard to exaggerate how important this is. Of course, any approach to philosophy of mind that isn't consistent with science is probably bankrupt. But any approach that actually extends the accomplishments of science by showing how they can be understood to solve fundamental issues has fulfilled one of the most important possibilities of philosophy, *real relevance to the discoveries of the day*.

C. Process/relationship relativity, learning and two leading research programs; connectionist architectures and computational modeling

The one thing we know is that brains are cognizers. At the same time we want to explore cognition with the most powerful and exacting methods at our disposal: Mathematical quantification. Connectionist description is highly exacting and mathematical. It is also remarkable in its basic resemblance to neurophysiological systems. So it collapses these two projects--the neurological and computational--very nicely into one. These are two very heady virtues in the latter Twentieth century of experimental psychology. For these reasons I will grasp the opportunity and explore process/relationship relativity in the connectionist context.

Because most of our contents are learned, the natural place to start in an explanation of error is in the context of "learning". "Learning" is a remarkably vague notion, but equally inescapable as a fact about most any cognitive system. Because learning is such an important source of contents, we need to ask what sort of processes will both generate contents and allow for misrepresentation via these? In what follows we'll focus on type-recognitions, like Rat, Rabbit or Spiny Hedgehog.

An opening suggestion,

> R has the content "Spiny Hedge Hog" relative to the processes of activating a bearer that was the subject of a previous process of, "successfully learning to recognize Spiny Hedge Hogs".

If such contents are learned there must be physical processes at work. What sorts?

An important first distinction is that of continuous versus discontinuous processes. Learning is often a discontinuous process, where several discreet episodes are necessary to form a pattern of contents that is well embedded in the larger cognitive system, or a motor skill or combinations of both. Learning at a single pass does happen, but as far as our basic perceptual range goes, it is uncommon. Our list of animals is built up slowly, and some of us took a year or two to realize that not every four footed furry is "kitty". So our story goes: R is embedded in a discontinuous learning process that fixes its content relative to that process. Later R is activated by a *not*-hedge hog; this is not part of the learning process R is embedded in, the process that discriminates spiny hedge hogs. Relative to the learning process R is embedded in, R has the false content "Spiny Hedge Hog".

Much more needs to said about this discontinuous learning process, since it is critical in providing R with a content that acts as a standard of correctness. In his early work on misrepresentation, Dretske offered this sort of theory to account for misrepresentation,

> Following the learning period L, the structure S acquires a life of its own....and is capable of conferring on subsequent tokens....its semantic content (the content it acquired during L) whether or not the subsequent tokens actually have this as their informational content....what this means, of course, is that subsequent tokens of this structure type can mean that [external object] s is F....despite the fact that they fail to carry this information....despite the fact that the s (which triggers their occurrence) is not F.[100]

That is, correct contents are fixed during a "learning period". *During* this period the content is fixed by actual samples (say, real horses) that are the objects of the content. Misrepresentation occurs when contents fixed during the learning period are activated *after* the learning period by objects that don't fall under the content (zebras at dusk).

---

[100] Fred Dretske, *Knowledge and the Flow of Information*, (Cambridge: MIT Press, 1981), p. 193.

But here the continuous nature of learning is the problem. The challenge to Dretske's account is simple: Suppose we are within a learning period for the concept **horse**, and a zebra activates this bearer; what sets this activation aside? If a zebra can activate the concept *after* the learning period, it can *during* it. So what makes the learning period a learning solely of **horse** and not **zebra** too? The disjunction problem appears again. The content learned is just as well **horse or zebra**. Our principled rejection of singular content allows us an easy answer, but this is not available to Dretske, because he accepts the choice problem. The worry for the physicalist (and others too) is that this talk of a learning period is *ad hoc* and simply imposed. There is nothing out there in the world that sets, external to the changes within the system, a standard period of time in which "learning" is happening. Learning is a particular kind of inner doing, it isn't just a time in an organism's life. But this is all Dretske offers.

Learning, internal changes and the micro-architecture of cognition; connectionist models

Our understanding of learning will have to include an attention to changes within the system. So we must look inside. The clearest, most successful accounts of learning *from the*

inside available today and the foreseeable future are *connectionist models.* These are the "artificial neural networks" that have surprised traditional theorists of learning and fascinated the world of artificial intelligence. We'll explore these next to get a better picture of the kind of processes process relativity can appeal to in its explanation of content and misrepresentation.

Supervised learning in connectionist architectures

"Supervised learning" in artificial neural networks is really a version of Dretske's learning period with an important twist: The network is supervised so that it's discriminations are *compared to* the correct results, and altered when they are off the mark. This is the way we might learn if we were under God's care as our ever attentive tutor. In such learning the network is fed a whole host of samples that are both of the type to be learned and not of this type. It is told, in effect, which is which. Acting on this information a learning rule continually alters the network so it will respond to the target types in one particular way, and to the nontarget types in another. After enough trials, the fundamental differences between the sample types is imparted to the network. Full learning is shown by the successful discrimination of *new* type and non-type samples.

But without the prior sorting of these during the learning period the network could not develop. That's why it is called *supervised* learning.[101]

For the most part, this kind of neural network cannot provide insight into human learning, for the same reason Dretske's learning period theory couldn't. Most of our learning isn't supervised. Without such supervision it's unclear how a stipulated "learning period" by itself could fix any determinate content relative to the processes occurring in that period. What about false positives during the period? What about false negatives? And even if the equivalent of supervision had occurred (there are, by immense luck, no false positives or negatives, and we set aside the worries about arbitrariness in the period's duration) there *might have been* false positives and negatives. Why think that because there wasn't this should have any effect on the content-fixing power of the period? Does the period just amount to this,

> Whatever single object-type activates R during period X is the content of R. X's duration varies according to when some new object-type activates R.

Or is an absolute value; seven days, months or years?

[101] Judith E. Dayhoff's *Neural Network Architectures* (New York: Van Nostrand Reinhold, 1990) gives an excellent and in depth overview of basic connectionist architectures.

What kind of physical fact could this correspond to? Either option is hopelessly arbitrary.

Learning can resume, be constant or fail. What decides which of these fates befall the period? In a nut shell: The period seems to have been endowed with special powers that *presuppose* a certain content to be fixed, and such a period, by itself, is incapable of doing this.

This may be taken to show learning periods are somehow suspect. But it is obvious that they do exist and play a critical role in our cognitive development. Most error exists relative to a history of learning. Can we capture this fact without imbuing learning periods with unearthly semantic powers?

Unsupervised learning in connectionist architectures

The learning processes we're interested in rely on a kind of principled self-supervision, on self-sorting powers of some kind. These enigmatic properties have been demonstrated in competitive network architectures like the *Kohonen feature map.*[102] The feature map is a powerful architecture that has these related properties,

1) The ability to find the organizational relationships among

[102] Ibid., pp. 163-188.

disparate input patterns.

2) Similarities among particular patterns of input are mapped onto closeness relationships on a competitive grid of neuron-like units.

3) So, after sufficient exposure to various input patterns, natural groupings and their relationships are observed in the organization of this competitive layer.

Sound familiar? Yes, these are many of the capacities that make human learning so powerful, self-monitoring and flexible. We can place the feature map in very different kinds of learning environments, and it will cope well, self organizing the distinctions its perceptual system encounters. Consider a *well-behaved* learning scenario like this one,

1) We have an environment with five different object types, I-V: Nazi Skinheads (I), neo-hippies (II), presidential candidates (III), corporate lawyers (IV) and Jesuit priests (V)....(one must really love this world), each of which can produce distinct inputs to our feature map, inputs that distinguish the types.

2) We build a perceptual system and hook it up to a Kohonen feature map. The map then self-organizes the objects' inputs into different topographical regions on the surface of its competitive layer. After enough presentations the network reaches a dynamic equilibrium from which it will not move. Learning is complete to the network's abilities.

Enter error: Imagine an input from a type I object activates a topographical region organized by inputs from type V objects. Question: What caused this error? Given the structure of the network and the context of its learning the possibilities are these,

a) The type V object's detectable surface properties are so similar to the type I properties that at this stage of self-organization, the map cannot distinguish them (this case of object type I "looks just like" an object type V case).

or,

b) The perception system, feature map system or both have changed in some way (perhaps temporary) so that, though in the past the type I object would have activated such and such region of the map, it now activates another.

or,

c) A refinement of (b): The relation of objects to detectable surface features develops further, embracing greater subtly. It *re*organizes itself in fairly extensive ways because of this. New feature discriminations add onto the old ones, preempting the old surface feature to object mappings.

or

d) A self mutating, continually fluid environment with no stable associations between surface features and object types.

The process/relationship approach to error interprets each of these cases easily.

Case (a): The pairs of contents in (a) are type I (skinheads) and type V (Jesuits). The story of (a): This kind of case is real-world enough. The hiker who accidentally stomps a toad into the great beyond says, "But it looked just like a rock to me....". The previous self-organizing period of the network is its learning history. Relative to this history, the learned association is between region I and skinheads, not Jesuits. If we are

interested in the previously learned contents being true, then relative to those learned contents, the activation of region I by Jesuits is an error. How did this calamity occur? A priest lost his hair and looked like a skinhead.

In case (b) the organization of the network has changed so that regions are activated by different types than before. Thus different processes now embed the representational tokens of map regions. These processes generate new contents for those regions. Contrasted to the old contents, they are in error. If we view the network as damaged, and are interested in the old contents, then the network is making mistakes.

In case (c) we encounter our cruel world, the very one young children find themselves in. This is mainly because a child's innate "preprocessing" in the perceptual system is so primitive and the child is continually making new identifying feature discriminations. So it is critical that we also address misrepresentation here. Consider vision: The network has learned to identify objects based on surface properties. The relation of objects to the surface properties are then changed as more surfaces properties are discriminated. New surface features, so new processes, are at work. If we take the old object/activation mapping as correct, the *new* will be the

erroneous mapping. If we take the new mappings as correct, the *old* will be the erroneous mapping. And a rat will no longer be **kitty**.

For the Kohonen feature map, as well as ourselves, (d) is the cognitive doom scenario: Learning reliable representation requires a certain stability in the environment. In an environment *continually* unkind in this respect, the network can neither self-organize to a stable state nor function from such a stable state.

Summing up; learning and error

There is a deep connection between learning and error--most all our errors are about learned contents, and many of our errors prompt more learning. The Kohonen feature maps explicitly displays both these properties. The map gives us a simple and clear way to view learning and its associated error. The map is subject to the most rigorous mathematical description of its development, inputs and self organization. Though we have described *learning periods* in cases (a-c), unlike Dretske we don't have any special explanatory dependence them, just on the organization of the network at given times and the different processes that regions of the network are associated with at these times. We have captured

the fact that learning periods are real aspects of a cognitive systems' development without imbuing these periods with magical powers.

Conclusion; process/relationship relativity a success

The results of this chapter are undivided good news for our approach. We have clearly distinguished it from the mainstream, showing how fundamental problems for the leading theories of content are easily solved by process/relationship relativity. The best news for any view is that it works and works easily. This appears to be the verdict on relativity. The ecumenical attitude and psychological ecology have paid off.

Chapter VI. Summing Up Psychological Ecology: Five Issues that Remain

So far

We have developed an approach to representational content that involves these ideas,

1) It is an attempt to give a generative/ontological explanation for content. The background ontology is "Naturalist". The general approach is captured by the “ecumenical attitude”.

2) It opts for an externalist solution to the source problem because of the explanatory opportunity (in the generative sense) found there.

3) It adopts the perspective of psychological ecology, one that views mental content as an expression of a cognitive system's representational tokens as situated in various processes and patterns of relations. These typically extend beyond the system and into its broader environment. This is the system's psychological ecology and it generates many of the system's contents. I have called this "process and relationship relativity" of content.

4) Psychological ecology and process/relationship relativity reject the choice problem--the problem of how to decide out of the myriad external relations that appear to generate content, how only one content for one token is generated (at least at any given time). Every representational token enjoys multiple contents, each of these relative to a different process or pattern of relations.

5) Psychological ecology and process/relationship relativity use multiple contents to solve the essential problems of any theory of content: The depth problem, disjunction problem, *qua* problem and the problem of misrepresentation.

6) Psychological ecology and process/relationship relativity cohere very well with connectionist models and computational models of cognition.

In broad outline I believe this approach is right. In many ways it is merely gestural though. We live in a time that is serious about applying science to the mind. If we embrace this ambition (I do) issues concerning its viability as science immediately confront psychological ecology. I have some

preliminary thoughts about this.

Psychological ecology also creates a powerful and disturbing question for our self knowledge: Do we know the contents of our thoughts? I will comment on this too.

Issue 1, A closer look at content assignment and processes

In the previous chapter we explored how process/relationship relativity, understood in an intuitive way, deals with the depth, *qua* and misrepresentation problems. Here we give a closer analysis of process/relationship relativity and the way it generates content assignments.

What is "process/relationship relative" content? *It is the hypothesis is that representational content is a matter of being part of a certain kind of physical process or patterns of physical relations. Different processes and physical relations provide different content. As we discriminate these, we discriminate contents.*

The method of assignments

Let's focus on *processes* here. I believe the notion of relations between tokens and the world is fairly straightforward, but what about getting content from process embedding? What is it about a process that specifies its particular content? My first

thought: It is the type of process that it is. Whatever this type relies on, this is the content. But what determines this type? Because processes consist of the entities involved and the changes these go through, we must look to these. On this view, content ascriptions are simply ways of describing physical process types and their parts.

There are a great many "parts" to a process, and which ones we attend to will determine the type we associate with a process. One of the most popular process parts to fix contents are the *limits of a causal chain.* That is, the beginning and endpoints of a chain. A simple, causal theory of external world perception is an example. "I see an apple" reports the process of a causal chain starting with the surface of an apple (the apple) and ending with the activation of a representational token in my brain (I see....).

Process relativity includes all physical causal relations, but to understand the full *range* of the typology of physical processes we must move beyond causal relations . We do this by acknowledging all the types a process falls under. At the same time content is virtually always embedded in the context of a psychological verb: Seeing, recalling, thinking, wondering, believing and so on. [79]

This psychological verb characterizes the

representational token and the cognitive side of the content generating process that token is involved in. Our specification of the process type needs to recognize this, and pair it off from the content generating, external world aspect of the process. More formally, the idea is this,

> *Representational token R is situated in a process, P. Process type P breaks down into two components; cognitive activity type A and an environmental type, E. Content is had by representational token R in A relative to process type E; R in A has the content* ***E****.*

A modest example,

> The visual cortex process that detects illumination discontinuities in my visual field has the content *edges in the environment* relative to a process that: (1) Starts with edges, includes photon fluxes that preserve the presence and orientation of these edges and arrives at the retinal surface. (2) has these photon fluxes transduced via the retina to neural patterns, patterns which are mapped onto aspects of the visual field that covary with these edges presence and edges distribution in the perceived environment. Given all this there is a process type, the *detection of edges in the environment.* These visual cortex mappings (a cognitive activity) reflect the *detection* of edges. Their content is **edges.....**

The key features of the process are causation by actual

edges and covariance in the visual cortex. Both causal and Information theorists will welcome this as an inclusion of their favored physical vectors under the process relativity motif. The process type consists of a cognitive activity component and an external world component. The causal and covariance relations of these fix the content of that process. Indeed, on the process relativity view, the "content" is just a shorthand for the edges in the world, their causal effect on the visual cortex, and the covariance this maintains.

Now let's generalize from this example. There are three steps that guide our content assignment,

(1) We outline the parts of the process to a degree of detail that is accessible and manageable.

(2) We distinguish the representational token and cognitive system aspects of the process (the act of "detection") from other aspects of the process (the "world").

(3) We assign a single content to the representational token based on cognitive system aspects that the representational token is embedded in (detection) and on the noncognitive aspects that define the content of the process type (the edges

in the environment).

One clarification: When the process is entirely bound up in the cognitive system (representation of the activities of one's own brain, for instance) then the cognitive and noncognitive aspect distinction is between those cognitive aspect that the representational token is embedded in and those it is not. (The limiting case is when the representation represents itself simultaneously, of course. Then the limits mutually collapse).

So the issue of what defines a "process type" is critical because this is what ultimately sets the content. A *process type* appears, in the case above, to be the "limits" of what it includes.The type is everything it includes to the limits of this inclusion. The limits give the type its name. In the edge example the limit of the process on the noncognitive side was *environmental edges*. This included many sub-processes, but the extreme was the edges themselves. Hence it was *edges* that were detected. Noncognitive processes appear to be specified by their limits too; burning wood is the process of rapid oxidation (burning) related to the thing being oxidized (wood). Many events--air flow, heat dissipation and conduction--and objects--oxygen, carbon, cell walls--fall within this, but the broadest characterizations--wood and burning, contain all the

rest and are not contained by something further *as far as burning wood goes*. They are the "limits" of the process.

It is true that there is some ambiguity in *exactly* where we put the limits of a process like "combustion". Events separated by nanoseconds may not be clearly distinguishable as falling within or without the process. This doesn't mean that these processes have no objective existence, however. At best it shows that they possess small regions of indeterminate boundaries. This seems the rule in the world, not the exception. Consider water: Putting spatial limits on the regions occupied by hydrogen atoms (as opposed to oxygen) in a water molecule is a task without an exact solution, but this hardly shows that there are not hydrogen atoms objectively present in water. What we stipulate as exact boundaries is of no concern--that there are causal chains of objective kinds (combustion) and that these have boundaries of either determinate or indeterminate (but nevertheless real) kinds is an objective fact, one that generates specific contents relative to the type of process involved.

This is a general formula. How far it can be extended, into how many contexts, is a matter for research. The theory of *process relative content* is just learning to walk. It is mature by no means. Only further exploration will determine how useful the approach can be.

Issue 2, Further Directions in Scientific Research

The Science dream is simple: We want a science of representation. The basic requirement for a theory of content that can be *exacting*, and not merely *gestural*, is that we be able to build mathematical models of the space of possible contents and a way to relate these to well described psychological ecologies, so that we can map contents onto ecologies. What counts as an adequate model and sufficiently "mathematical" are critical and very involved questions. Understanding how psychological ecology works within this task is an important issue.

But prior to these questions we need some idea how we might proceed in *any reasonable sense* to use quantified analysis in representational science. Is there any opportunity here? Are there any directions we might take? The broad (and admittedly vague) task is uncontroversial: We need to mathematically map a cognitive system's psychological ecology to the contents this generates. The variables appear to be straight forward,

1) representational tokens

2) processes and relations they are embedded in

We must carefully characterize these. Understanding the multitude of ways to go about this is an important direction of

further research. Understanding (1) is a burden all generative theories of representation bear. (2) is a task for process relativity, but versions of it are involved in causal theories of perceptual content too. What are the apparent opportunities for understanding of these?

The most direct way to understand (1) is in terms of physical objects and activities: Neural arrays, patterns of activation and depolarization, computational hardware and its electrical behavior and other physical structures and activities. A highly developed taxonomy and quantification already exists for these and their properties. (2) is an empirical issue too: The best way to approach processes is to determine what behavior the organism is undergoing that interests us--for instance, we seem very interested in survival behavior, behavior characterized in terms of other, more subtle end states and communicative action--and ask what processes the system is embedded in that allow us to characterize the generation of that behavior representationally.

We can go about this two ways. We catalog the processes apparent in the system's ecology and then relate these to representational tokens, and then relate these to the behavior, or postulate a system of processes and representations that lead to the behavior, and then try to

discover these in the system and its ecology.

But how could these variables be quantified? The approach is simple in concept: Take measurable environmental variables and measurable behavior and develop a formula that (a) transforms the environmental quantities into the behavioral quantities and (b) have the mechanisms in this transformation be entirely justified by the facts of the animal and its environment--no arbitrary constants or other mechanisms that cannot be linked to the animal or its relation to the environment. Perhaps this can be extended to the human animal and its environment.

Issue 3, Contents and cognition

There is a further issue that is very important: The transformation of contents into descendent, but different contents. Essentially, the act of *thinking.* This ability is the holy grail of cognition and *the* key area for research on *cognition.* Process relativity accounts for initial contents, but the work of transforming these into trains of thought is a very different task. Once we fix initial contents in a cognitive system we need to ask, "what next"? We need to understand how physical processes within the cognitive system generate changes in the contents of the system's representational tokens.

Here the background perspective of psychological ecology reveals its deeper nature. What are we characterizing when we give an ecological account? We are characterizing the ecological aspects of a cognitive system's *states and activities*. This reflects on a pattern of changes within and without the system, and so ultimately it is *causal processes* we are characterizing. These causal processes issue forth new representational tokens with contents that are modified relative to earlier stages of the processes.

Some necessarily vague ideas on how this might occur,

1) Simple combining, where tokens are united into complex wholes by causal processes.
2) Causal activity might take previously complex tokens and deactivate portions of them, eliminating the presence of that content, pairing down the complex representation into simpler parts.
3) The flip side of (1), tokens can be replicated to build up aspects of a representation--like magnitudes, or subtypes, so that the content is skewed in some ways.
4) Resemblance transformations might also play a role in producing new contents. By resembling a token that resembles one with certain contents, we can chain the content into new

tokens. Such chains can pair down and combine contents. Causal changes to a token might alter the resemblance transformations that it is involved in, modifying its resemblance contents.

These suggestions are by no means intended to be the foundation of an account of cognitive activity. They may not prove ultimately viable. They merely gesture to the sort of things we might consider within the approach of psychological ecology.

Issue 4, Psychological ecology and our knowledge of contents

Sometimes views are valuable as much for the questions they lead to as they are for the answers these questions grew out of. Our notion that we are "directly aware" of the contents of our thoughts, and how externalism must deal with this is a case. Externalism is certainly one case. Consider this simple argument,

1) We are directly aware of the content of our thoughts in the very act of experiencing those thoughts.

Seemingly a give away, this is just the idea that when I think, "that rabbit is brown" I understand what I'm thinking--I understand what a rabbit is, what brown is and how a rabbit--

her fur--can be brown. And all that immediately confronts my awareness when I think what we would mark in English as the thought, "That rabbit is brown." What makes this understanding *direct* is the idea that it is not the result of any inferences, evidential weighing or other kinds of informed speculation. It's existence simply is given by the fact the thought occurs to us.

2) We are not directly aware of the external relations and process that according to process/relationship relativity (or any externalist theory) generate our contents.

This means that we don't directly know what our world is like. Our knowledge of the world is at best evidential and articulated with endless assumptions and cognitive contrivance, none of which are certain, let alone direct. But therein is the mystery. If we aren't directly aware of these processes and relations, how can we be directly aware of the contents they generate? For all we know the world--and so the contents it delivers--is very different than we believe it is. If we have direct access to anything, it is simply the activities within our brain, and only a few of those. But these under determine what their contents are, they can remain unchanged even though different psychological ecologies would invest them with different contents. So how can we possibly be *directly aware* of these

contents? In a premise,

3) In externalism, what we are directly aware of under determines the content of any representation.

4) So if externalism was true, we would not be directly aware of the contents of our thoughts.

Recalling premise (1),

5) So externalism is false.

What's wrong with this argument? Without getting into subtleties that miss its basic force, the problem is premise (1). We are *not* directly aware of the contents of our thoughts. As one of the basic tasks in clarifying psychological ecology, we need to explain why and characterize what we are aware of--since we are certainly aware of *something*.

There are many things to be said here.[103] The most direct is to provide an alternative description of what we are aware of when we experience our "thoughts". This needn't be contents. In the human case what we distinguish and have experiential access to is brain processes and features of their structure.

[103] See Tyler Burge, "Individualism and Self-Knowledge", *Journal of Philosophy*, Vol. 84 (1988): 649-663. Though fascinating, I believe a more direct approach is advised.

What we know when we know what we are thinking is this, and with moment to moment exploration of this, some of the connections and relations these have to other brain processes and structures. Chief among these is how these translate into natural language sentences. I know what I am saying is akin to knowing what I am thinking. Indeed, there is little difference. And experiential knowledge stops there. Facts about what our linguistic expressions ultimately mean are no more immediate than facts about the contents of our thoughts.

Agreed, this needs refinement, further explanation and further justification. But it is a beginning on a powerful view of the very nature of our self-concept.

Issue 5, a last thought on the nature of content and our content ascriptions

The final question is huge: What has content become in psychological ecology? How far have we come from the crude *syndrome* outlined in chapter one? I will only offer a few words.

We have attempted a real understanding of this syndrome. In doing so content has become something that is *immense,* something which *totally intertwines us with the world.* Whether we know it or not, externalism diffuses and intertwines our minds with the very nature of our world. Psychological

ecology and process/relationship relativity takes this intertwining to an entirely new level. Our mentality becomes in its representation infinite in extent, reflecting virtually every facet of the universe. We are total in our representational relationship to the world, the ultimate nexus of it and all its aspects. Our cognition is the difficult, continuing discovery of this fact. That, to my mind, is *beautiful.*

BIBLIOGRAPHY

Adams, Fred. "The Rutgers Chainsaw Massacre IV: Jason Goes Sailing." Unpublished draft.

Adams, Fred and Aizawa, Kenneth. " 'X' means X: Semantics Fodor-Style.", *Minds and Machines,* **2**, (1992), pp. 175-183.

Berkeley, George. *A Treatise Concerning Human Understanding.* Reprinted in *George Berkeley: Principles, Dialogues, and Philosophical Correspondence.* Colin Murray Turbayne (ed.). New York: MacMillan Publishing Company, 1965.

Burge, Tyler. "Individualism and Self-Knowlegde.", *Journal of Philosophy,* Vol. 84, (1988), 649-663.

Churchland, Paul. "Conceptual Diversity Across Sensory and Neural Diversity: The Fodor/Lepore Challenge Answered." *Journal of Philosophy,* Vol. 95, (1998), pp. 5-32.

Clark, Austen. "Mice, Shrews and Misrepresentation." *Journal of Philosophy,* (1993), pp. 290-310.

Cummins, Robert. *Meaning and Mental Representation.* Cambridge: MIT Press, 1987.

_____. *Representations, Targets and Attitudes.* Cambridge: MIT Press, 1996

Davidson, D. "Radical Interpretation," *Dialectica,* 27, pp. 313-328.

Dayhoff, Judith E. *Neural Network Architectures.* New York: Van Nostrand Reinhold, 1990.

Dennett, Daniel C. *The Intentional Stance.* Cambridge: MIT Press, 1987).

Devitt, Michael and Sterelny, Kim. *Language and Reality: An Introduction to the Philosophy of Language.* Cambridge: MIT Press, 1987.

Dienes & Berry. "Implicit Learning: Below the Subjective Threshold," Psychonomic Bulletin & Review, 1997, 4(1), pp. 3-23.

Dretske, Fred. *Knowledge and the Flow of Information.* Cambridge: MIT Press, 1981.

_____. "Misrepresentation." In Bodgan, R (ed.) *Belief: Form, Content and Function.* Oxford: Clarendon Press, 1986.

Fodor, Jerry. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind.* Cambridge: MIT Press, 1987.

_____. *A Theory of Content and Other Essays.* Cambridge: MIT Press, 1990.

Godfrey-Smith, P. "Misrepresentation." *Canadian Journal of Philosophy, 19,* pp. 533-550.

Haldane, John. "Putnam on Intentionality." *Philosophy and Phenomenological Research,* Vol. LII, No. 3, (1992).

Hannan, Barbara. *Subjectivity and Reduction: An Introduction to the Mind-Body Problem.* Boulder: Westview Press, 1994.

Kripke, Saul. *Naming and Necessity.* Cambridge: Harvard University Press, 1980.

Laurence, Stephen and Stich, Stephen, "Intentionality and Naturalism", reprinted in *Deconstructing the Mind,* (Oxford: Oxford University Press, 1996), pp.168-191.

Lycan, William. *Logical Form in Natural Language.* Cambridge: MIT Press, 1986.

Millikan, R. G. "Biosemantics" *Journal of Philosophy* 86, pp. 281-97.

_____. *Language, Thought and Other Biological Categories* (Cambridge:MIT Press, 1984).

Plato. *The Republic.* Desmond Lee (trans.). New York: Penguin Books, 1974.

Papineau, David. "Representation and Explanation."*Philosophy of Science* 5, pp. 550-72

Putnam, Hilary. "The Meaning of 'Meaning'" in Putnam, Hilary, *Mind, Language, and Reality: Philosophical Papers,* Vol. 2, pp. 215-271. New York: Cambridge University Press.

_____. *Representation and Reality.* Cambridge: MIT Press, 1988.

Quine, W. V. O. *Word and Object.* Cambridge: MIT Press, 1960.

Reid, Thomas. *Inquiries and Essays,* (eds.) R. E. Beanblossom and K. Lehrer. Indianapolis: Hacket, 1983.

Schroeder, Manfred. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise.* New York: W. H. Freeman and Company, 1991.

Searle, John. *Intentionality.* Cambridge: Cambridge University Press, 1983.

_____. "Minds, Brains and Programs." *Behavior and Brain Sciences 3* (1980), pp. 417-458.

_____. *The Rediscovery of the Mind.* Cambridge: MIT Press, 1992.

Shepard, G. M. *Neurobiology.* New York: Oxford University Press, 1988.

Sterelny, Kim. *The Representational Theory of Mind.* Cambridge: MIT Press, 1991.

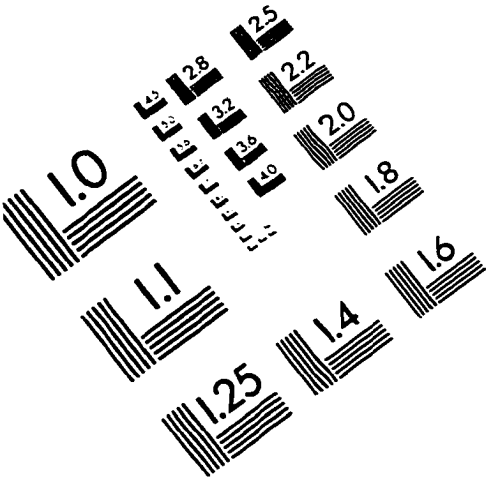
Stich, Stephen. "Dennett on Intentional Systems," *Philosophical Topics,* 12, pp. 38-62.

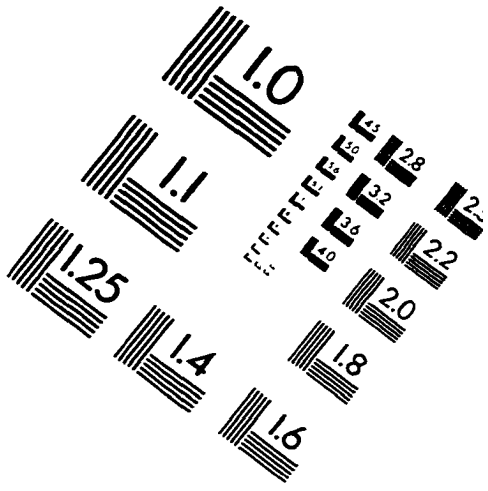
_____. "Do Animals Have Beliefs?", *Australian Journal of Philosophy,* Vol. 51, (1979).

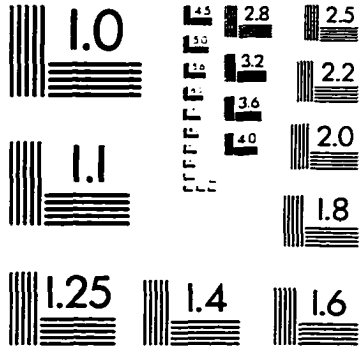
_____. *From Folk Psychology to Cognitive Science.* Cambridge: MIT Press, 1983.

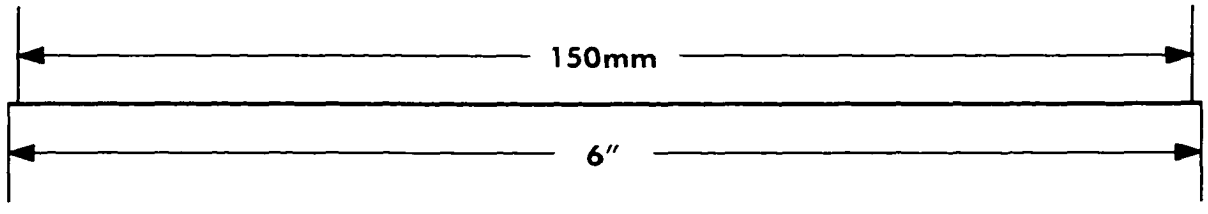
Tye, Michael. "Naturalism and the Mental," *Mind,* Vol. 101, No. 403 (1992). pp. 421-439.

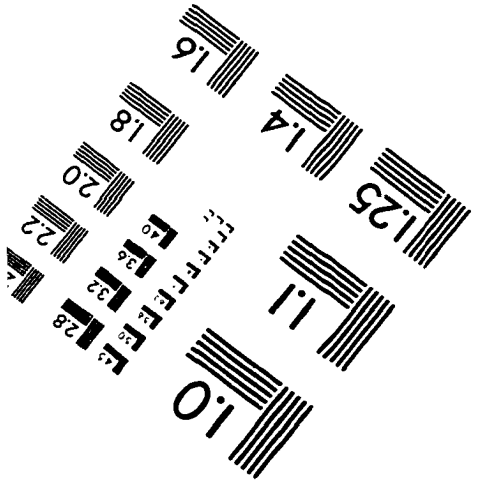
IMAGE EVALUATION
TEST TARGET (QA-3)

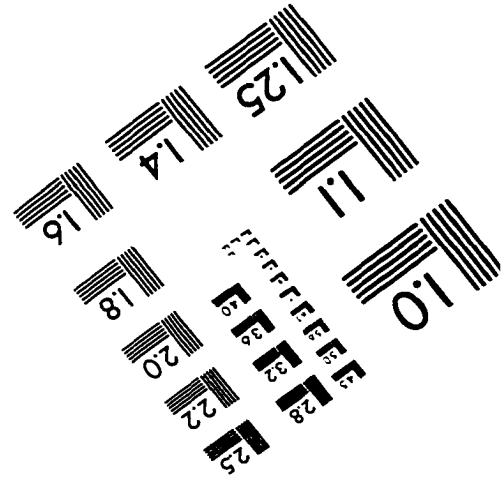












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989