Chapter 3
More about Error

Forced Error and Expressive Adequacy

In this section, I distinguish two kinds of representational errors: "forced errors," which are unavoidable given the expressive power of the representational scheme, and "unforced errors," which are not due to any inadequacy in the scheme itself. Let's begin with some examples.

- No finite language, that is, no language with only finitely many expressions, is adequate to represent the facts of arithmetic.
- The Fahrenheit scale, being an interval scale, is inadequate to represent certain facts about temperature, namely ratios.
- Extensional logics appear to be inadequate to represent the logical form of subjunctive conditionals.
- Euclidean geometry has proved inadequate to represent the structure of physical space.
- People seem unable to see a human nose as concave rather than convex.
- Three-color schemes are not adequate to represent all possible territorial boundaries.

These are all cases of expressive inadequacy: a certain representational system just does not have the power to represent a certain domain. One of the things that makes such cases interesting is that they generate situations in which there is no right answer to the question of how to represent a given target. In general, there will not always be a best answer, or even a reasonably small set of best answers, and even when there is a best answer, it might not be very good. Borrowing a term from tennis, I call misrepresentations that arise from the expressive poverty of the representational scheme *forced errors*: forced errors are forced in that there is no possibility of a correct representation in such situations. The representational scheme simply lacks the requisite expressive power.

A central point about expressive adequacy is that it is a relation between a representational scheme and its target domain: an expressively adequate scheme is one that has the resources to "adequately represent" the domain that is its target. The relevant notion of adequacy here probably cannot be made precise without sacrificing generality. Different kinds of schemes and domains may well call for different conceptions of adequacy: the inadequacy of extensional logic (if such it is) to represent the logical form of subjunctive conditionals is not obviously of a piece with the inadequacy of three-color schemes to represent boundaries. Haugeland (1990) may well be right in suggesting that different kinds of representational schemes represent different kinds of things. For example, it may be that there are things that can be represented iconically that cannot be represented symbolically and vice versa. If so, then there will be things a representational scheme cannot represent just because it is the kind of scheme it is. There is rich material for research here. Still, we seem to have a pretty good intuitive grasp of the idea that a representational scheme might or might not be up to the job of representing some specified domain, and so I have little hesitation about taking this idea as primitive in the following discussion.

We have to be careful to distinguish the expressive power of a scheme of representation from the capacity of a particular system to use that scheme. English grammar allows for unlimited center-embedding, but human parsing architecture blows up at a depth of about 3, apparently because of limitations on short term memory (Miller, 1956). Paul Smolensky (1987) has developed a scheme of distributed representation in connectionist networks that allows for the representation of complex hierarchical structures such as phrase markers, but until very recently (Smolensky, LeGendre, and Miyata, 1992) it was an open question to what extent connectionist processors could exploit the representational power thus achieved (Smolensky, 1994; Cummins, 1994). One of Galileo's most original ideas was to make the areas of Euclidian figures represent distances. Letting vertical lengths represent time, and horizontal lengths represent velocity, the area of a rectangle represents the distance traveled by a body moving with uniform velocity, and the area of a right triangle represents the distance traveled by a uniformly accelerated body. When acceleration isn't uniform, we get a right "triangle" with a curved hypotenuse. Here, the area represents the distance, but the standard straightedge and compass geometry of the day had no means of comparing the resulting area with others, and was therefore unable to exploit these representations.[1]

Forced error, then, arises when a system Σ uses a scheme R for representing a domain T that R is not capable of representing. The situation is this: Σ makes a forced error in representing t as r when

1. The right curves could be produced accurately on the hypotenuse sides by tracing a conic section.

1. R is Σ's scheme for representing T.
2. Σ applies r (an element of R) to t.[2]
3. No element of R represents t.

At the point in the process where a representation of t is required, Σ is bound to use some element of R. But no element of R represents t, and so Σ's use of r to represent t is a forced error.

Not all representational error is forced, of course. People frequently misrecognize shrews and voles as mice. Sometimes, this is because they don't know about voles and shrews, so some of these errors might be forced errors. But even a zoologist who knows about voles and shrews might mistake a vole for a mouse on occasion, and this is an unforced error in that it is possible, so far as representational resources go, to avoid misrepresentation.[3]

Unforced errors are common. Logic students, asked to represent the logical form of, "I won't shoot unless he does," typically suggest "$H \supset I$" rather than "$I \supset H$" which is also in their repertoire. Illusions are typically unforced errors: It is not because we lack the representational resources to see it smaller that we see the full moon near the horizon much enlarged, nor does an afterimage, seen as a spot on a wall, appear to shrink when we approach the wall because we lack the capacity to represent the spot as constant in size (as the phenomenon of size constancy proves).

Unforced errors, then, differ from forced errors in that the scheme R that Σ uses to represent T actually can represent what needs to be represented. We get unforced representational error when Σ selects the wrong element of an adequate scheme.

1. Σ's scheme for representing T is R.
2. Σ applies r (an element of R) to t.
3. R contains an element that represents t, but r does not represent t.

Interesting empirical and theoretical questions can be seen to turn on whether certain errors are forced or unforced, or due to an inability to exploit an adequate scheme.

- Piagetians think a first-grader's errors on word arithmetic problems are forced errors; their opponents think they are unforced, due to mistakes in processing, not to expressive poverty (D. Cummins, 1988).

2. By an element of R, I mean a representation, whether simple or complex, that is part of the scheme R, a representation that is, as it were, well formed in R. This might be a picture, a diagram, a graph, a symbolic expression, an activation pattern, etc., depending on the scheme.
3. Even the zoologically untutored can have percepts of voles and shrews, of course, just as cameras can produce pictures that represent voles and shrews without having the corresponding concepts. But the capacity to perceptually represent a vole or shrew must be distinguished from the capacity to represent the property of being a vole or the property of being a shrew. Percepts, as Kant realized, are pointed at the *objects*, not the properties.

- Barring encapsulation, effects resting on unforced errors should be cognitively penetrable, whereas those resting on expressive inadequacy should not be.
- Learning, in the sense of improved performance, divides into three distinct kinds: (1) improvements in expressive power, (2) improvements in the capacity to exploit existing expressive power, and (3) improvements in inference (broadly construed) that reduce unforced errors.
- Expressively powerful schemes may be difficult to exploit, but weaker schemes may lead to forced errors which are difficult to correct because a new scheme amounts to an alteration in architecture.

Doubtless other issues would benefit from reconsideration in light of the distinction between forced and unforced errors, and the related distinction between the expressive power of a representational scheme and the power of a computational system to exploit it. But let us leave empirical theory to those whose errors are less likely to be forced, and focus instead on the implications for the theory of content.

Accuracy: Degrees of Correctness

Representational error and correctness often come in degrees. It is obvious on the face of it that some pictures, maps, and graphs are more accurate than others. For this reason, it is desirable to replace the categorical notion of correctness with the graded notion of *accuracy*.

Note that this move makes good sense only if we are careful to distinguish accuracy from truth and inaccuracy from falsehood. Graded notions of truth and falsehood are notoriously difficult to make coherent, yet everyone knows that some sentences, maps, graphs, and pictures are more accurate than others. We speak of getting closer to (or further from) the truth, yet this tends to degenerate into getting more (or fewer) truths in the face of the fact that a failure to express a truth is simply a falsehood (or nonsense). The solution to this paradox is to note that accuracy is a relation between representation and target, whereas truth is a relation between representation (or attitude) and fact. A representation r may be closer to the truth than r' because it is a more accurate representation of some truth than r'. This doesn't make r "more true" than r'; it simply makes it a more accurate representation of a target that happens to be a true proposition. I'll have more to say about accuracy in chapter 7. For now, I will simply take it for granted that representations can more or less accurately represent their targets.

Seriousness and Inaccuracy

The degree of error or inaccuracy should not be confused with its *seriousness*. The *seriousness* of an error depends on the extent to which the outcome of a process depends on accuracy. Representing mass as weight can give large errors that are not serious in certain cases, for example, in processes designed to determine only which of two nearby masses is the greater. On the other hand, very small errors can be very serious, as the annual death toll from mushroom poisoning testifies. Once we see that an error can be small but serious, or large but not serious, we see that what we might call the *effectiveness* of a representation (on a particular occasion or generally) is distinct from its accuracy. Even a perfectly accurate representation can fail to be effective because it is too costly to use.[4] Given limited space, a map that leaves out many features and distorts others may be more effective than a more accurate map that will fit in the available space (a single atlas page, say) only at the price of being unreadable. Approximations to a target (e.g., linear approximations to nonlinear functions) are often more effective because more tractable.

Given that it is, in a sense, effectiveness that matters, it is tempting to define correctness or accuracy in terms of effectiveness: A correct representation is one that is effective. This is the intuition that helps to drive conceptual role semantics (CRS) as well as adaptational role theories. But this intuition must be resisted. If we define accuracy in terms of effectiveness, we are left with no natural way to articulate the important point that the pursuit of accuracy is often too expensive to be an effective policy. This is just one aspect of a central theme of this book, viz. that use theories undermine the notion of error, and hence undermine the explanatory value of the concept of representational content. We require a clean distinction between accuracy and effectiveness if the concept of representation is to have the independent explanatory leverage that makes it an important primitive in cognitive science.

4. We must not confuse a situation in which a system is effective in spite of its errors with a system that makes effective errors.

Less accurate representations are often tolerable because they are less costly to compute. Misrepresenting a crow as a hawk is a less serious error for a field mouse than misrepresenting a hawk as a crow. Given that recognition must occur quickly, a fast but inaccurate system may be better than a slower more accurate one. Since crows greatly outnumber hawks, a fast system that generates many false-positive hawk identifications but no false negatives is less accurate but more effective than a slower system that generates fewer false positives while still avoiding false negatives.

Note, however, that this is not a remark about the relative effectiveness of crow and hawk representations, but of the relative effectiveness of two different recognitional systems. Given the presence of a crow, there is no reason to suppose that a |hawk| is the more effective representation. On the contrary: It will cause the mouse to lose time and energy.

Chapter 4
Use and Error

A central theme of this book is that the theory of representational content must abandon any form of the doctrine that meaning is use. I think *all* of the theories currently taken seriously are use theories of one kind or another. The only major contender that is self-consciously a use theory, however, is conceptual role semantics (CRS). If we want to see what is wrong with the idea that meaning is use, therefore, we do well to begin with CRS.

A word about "use". To use a representation is to apply it to a target. Uses, then, are simply applications. To specify how a representation is used on a particular occasion is to specify a particular target. To specify a general use of *r* ("how *r* is used") is to specify what targets *r* is (or can be) applied to. Use theories of meaning (representational content), then, take the meaning of a representation to be determined by what it is or can be applied to. The fundamental idea is simple: in a case of correct use, content = target. So if we know what a representation is applied to, and we know that the use is correct, we know the content. Use theories attempt to naturalize content by specifying a sufficient naturalistic condition under which use can be supposed to be correct. They then identify the content of *r* with whatever *r* is applied to when the specified correct use condition is satisfied. In chapter 5, we will see that causal theories are use theories in this sense, for they identify the content of *r* with its target when it is applied by a successful detector. Adaptational role theories are also use theories, for they identify the content of *r* with the target it was applied to in those cases that account for the replication of the applying mechanism (intender).[1]

1. As we will see in chapter 7, there is a "narrow" and a "wide" aspect to target fixation. We have, for example, an intender whose business it is to represent *the current board position* (narrow). But what the current board position happens to be on a given occasion will, of course, be determined by the world. A use theory of meaning gives you narrow contents if it identifies contents with (correctly hit) targets identified in terms of intender function; you get wide contents if it identifies contents with the (correctly hit) targets themselves.

It is a little trickier to see how CRS fits this template. CRS is what you get when you put functionalism about mental states together with the idea that some mental states are individuated by their contents: If functional role is what distinguishes mental states, and if believing that p is distinct from believing that q, then the difference between believing that p and believing that q must be a difference in functional role. Having come this far, the only issue remaining is what sorts of functional roles fix meaning.

The idea is to think of the meaning of a representation in a system Σ as fixed by the cognitive transitions it enables in Σ. This is analogous to the line empiricists took about the meaning of theoretical terms in science (Hempel, 1950). Just as empiricists wanted to identify the meaning of a theoretical term with the conceptual role it plays in its home theory, so CRS identifies the meaning of a mental representation with the conceptual role it plays in a kind of "automatic theory," namely a cognitive system.[2] We think of the functional roles that individuate contents as *conceptual* roles because we think of concepts as the things that make it possible for cognitive business to go forward. You cannot infer that McKinley is dead from the fact that he was assassinated if you don't have the concept of death. But if you cannot infer McKinley's death from his assassination, you do not seem to have the concept of assassination either. That, at any rate, is the intuition underlying CRS (Stich, 1983).

It is now pretty straightforward to see CRS as a use theory in the sense lately staked out. Use theories say that the meaning or r is determined by what Σ applies it to. What Σ applies r to will turn on what Σ believes (and its other attitudes), together with the stimuli that impinge on it and its computational architecture. These factors are just r's conceptual role in Σ. So conceptual role determines use, and use determines meaning.

CRS and Meaning-Incomparability

The Standard Objection to CRS is that systems that differ in their beliefs, or any other of their attitudes, will be *meaning-incomparable*: no mental

2. This, in fact, is just how computational theories of cognition were originally conceived. Linguistic competence, for instance, was to be explained on the hypothesis (1) that one had tacit knowledge of linguistic theory, and (2) that an automatic inference and control system applied the theory to produce and understand speech. If you hold a conceptual role theory of the meanings of the theoretical terms in linguistic theory, then it is natural to hold the same theory about the meanings of the terms of the tacit theory.

This little history lesson should give us pause. Conceptual role stories about the meanings of theoretical terms in science have been generally abandoned. It is hard to believe that what made CRS bad philosophy of science will somehow go away if we just put the theories in our heads.

representation in the one will mean the same as any mental representation in the other. Hence, systems that do not share *all* of their beliefs do not share *any*.

Here is how it is supposed to work. According to CRS, the meaning of a representation in a system Σ is determined by its conceptual role in Σ. One can picture conceptual roles (or functional roles, if you prefer) as follows. Each belief, desire, and so on, is a node in a connected graph, each connection representing what Fodor (1987) calls an epistemic liaison. The content of a given node is specified by its place in the rest of the graph. Adding or subtracting a node from the graph thus changes the meaning of every node. So, as threatened, systems that do not share *all* their beliefs do not share any.

An alternative route to the same conclusion goes like this. The content of a belief is given by its epistemic liaisons, that is, by all the evidential relations it bears to other attitudes (and perhaps other nonattitudinal states as well). Now, as Quine emphasized (1953), anything can be evidentially related to anything else via some connecting background theory. Hence, any difference in background theory will produce a difference in the epistemic liaisons that characterize a given belief.

It is generally thought that CRS can avoid this consequence only by invoking an analytic/synthetic distinction. If we allow only analytic connections in the graph that determines meaning, then meaning will be insensitive to mere differences in "collateral information" (Quine, 1960), for those differences do not effect analytic connections. There is, to be sure, a widespread intuition to the effect that one can have beliefs about arthritis even though one has some beliefs that turn out to be analytically false when construed as about arthritis (Burge, 1979). This intuition can certainly be challenged (Fodor, 1982), so the determined defender of CRS might feel reasonably safe if only a defensible analytic/synthetic distinction were on the cards. It is pretty widely conceded, however, that a defensible analytic/synthetic distinction is not on the cards, so CRS appears to be stuck with meaning-incomparability across persons and times. This line of argument is a central theme in Fodor and Lepore (1992).

It turns out, however, to be fairly easy to formulate CRS in a way that avoids the charge of meaning-incomparability. The new formulation does bring a kind of analytic/synthetic distinction in its train, but it is no more than what is already implicit in any computationalist theory that accepts the distinction between a system's attitudes on the one hand, and its fixed functional architecture on the other (Pylyshyn, 1984). It is arguable that connectionists do not have to accept this distinction (Schwarz, 1992), but few of those who have charged CRS with meaning-incomparability will be willing to take out connectionist insurance policies.

A Functionalist Version of CRS

The idea underlying CRS is that the meaning of a representation in a cognitive system is fixed by all the epistemic or evidential connections it enables, that is, by what Fodor calls its epistemic liaisons. The first thing to note when one comes to think about implementing this idea is that representations themselves do not enter into epistemic liaisons. Beliefs or desires or intentions can be justified or rational or warranted or probable on the evidence, and they can enter into making other propositional attitudes justified or warranted or rational or probable on the evidence. But a representation by itself can do none of these things. A representation, by itself, is *semantically* assessable but not epistemically assessable. My belief that war is hell may be justified by other beliefs, or by perceptual inputs, but my representation of war as hell, while it might be more or less adequate or accurate or useful, is not the sort of thing that could be justified or warranted.[3] Its contribution to the epistemic economy depends on what attitudes it enters into. A representation with the content *that I am rich* may enter into a rational desire or intention but an irrational belief. Representations thus determine epistemic liaisons only indirectly, via the attitudes they figure in.[4]

It is common, therefore, to find CRS formulated and discussed in a way that implies that it is attitudes, not representations, whose semantic content is determined by their epistemic liaisons. The initial idea is simply to identify an attitude's content with its epistemic liaisons. This view will entail that a belief *that p* and a desire *that p* do not have the same content, since a belief and a desire will never have the same epistemic liaisons. Indeed, they will typically have incompatible epistemic liaisons, since the desire that *p* typically has as an epistemic liaison the belief that not-*p*. This is hardly an unexpected consequence, though it is certainly not one advocates of CRS typically notice, let alone rush to embrace. If we want a CRS account of what it is for two attitudes to both be attitudes *that p*, we will require something more sophisticated than simply identifying content with epistemic liaisons. I propose to let this pass, since my real concern is with representational content, not attitude content.

The Formulation

Since representations affect epistemic liaisons only indirectly via the applications, and, through them, the attitudes they make possible, we need

3. Another way to put this is to say that only the *use* of a representation can be justified or warranted. This is why epistemological theories of meaning are inevitably use theories.
4. The same goes for applications. The same application can enter into different attitudes. Since instances of different attitude types are always characterized by different epistemic liaisons, it follows that epistemic liaisons alone will not give us the contents of applications.

to begin with a clear conception of the relation between representations and attitudes. Following Schiffer (1987) let's think of each attitude type as a kind of box—a belief box, a goal box, etc.—into which the system can put various applications. To get the belief that I am rich, my cognitive system puts an application with the content *that I am rich* in my belief box; to get the desire that I am rich, my cognitive system puts an application with the content *that I am rich* in my desire box.[5] I don't know how many kinds of boxes there are in a human cognitive system, or what they are, but that won't matter, because I propose to quantify over them in what follows. All that matters, then, is that there is a principled distinction between the representations a system uses, the representational uses it puts them to, and the cognitive functions of those uses.

Since it is attitudes that, in the first instance, have epistemic liaisons, let us begin by articulating a precise representation of the epistemic liaisons belonging to a given attitude. Begin with an A-state c_A, that is, cognitive state c that contains an attitude A. The set of paths in state space that intersect at c_A are the epistemic liaisons of A in the context of the state c_A, or A's conceptual role relative to c_A. But, of course, A can occur in others states as well. The conceptual role of A is just the set of A's relative conceptual roles, that is, the set of conceptual roles relative to c for each state c in which A can occur. According to CRS, the content of A is, or is determined by, its conceptual role.

You can get an intuitive feel for this definition by thinking about the consequences in a system Σ of the occurrence of an attitude A. A, of course, has no consequences by itself, but only in the context of a complete state of Σ. So think of completing A somehow, by adding other attitudes and whatever else it takes to make a complete state of Σ. All the paths in Σ's state space that intersect at this state are then all the ways the system can develop from A in the context of that completion, together with all the histories that could lead to A in that completion. A's total conceptual role is just the collection of all its roles relative to some particular completion.[6]

Can we now parlay this account of attitude content into an account of representational content? An attitude has definite effects only in the

5. I am not quite following Schiffer here, of course, for Schiffer supposes that to get belief that I am rich you put a representation with the content that I am rich in the belief box, rather than putting an application with that content in the belief box. See chapter 2 for a discussion of the distinction between applications and representations.

6. Note that CRS cannot distinguish the contents of attitudes that always co-occur in Σ. Note also that we cannot just take the union of the relative conceptual roles, for then we will be unable to distinguish two attitudes whose relative conceptual roles are the same because for every completion of one there is a different completion of the other that yields the same path.

context of a complete cognitive state. *A* representation has effects by participating in an attitude, that is, by being applied to a target and put in a box (given a semantic and cognitive role). *A* representation must inherit its epistemic liaisons from the attitudes it makes possible, so, having already defined the conceptual role of an attitude, we can identify the conceptual role of a representation with the set of attitudes it enables. Alternatively, we can proceed as before: Let *c* be some state containing *r*. The conceptual role of *r* relative to *c* will be the set of paths intersecting at *c*, and the conceptual role of *r* will be the set of *r*'s relative conceptual roles.

These definitions allow us to capture precisely the contribution a given representation makes to a cognitive system. Its meaning is identified with the paths through Σ's cognitive space that are made possible by the availability of that representation in the system.[7] Notice, however, that the set of epistemic liaisons determined by a given representation is completely insensitive to which beliefs (or other attitudes) a system *actually* has. It thus does not have the consequence that a failure to share beliefs entails a failure to share meanings.

Why is it generally supposed that CRS inevitably makes what one means depend on what one believes? It arises as follows. One begins with the claim that meaning depends on inferential connections. What one can infer from *P*, however, depends on what else one believes, unless we count only the strictly analytic consequences of *P* by itself. We can't do that, however, because there is no analytic/synthetic distinction.[8] So, what is inferable from *P* depends on what else one believes. So, if meaning depends on inferential connections, what *P* means depends on what else one believes.

This perspective—the perspective of Σ's current epistemic situation—is appropriate to epistemology, but it isn't appropriate to semantics unless one is a radical verificationist. What I can now infer from *P*, given what I now believe, might be thought to determine *P*'s epistemic significance for me now. But *P*'s meaning is more plausibly identified with what I could infer from it given a variety of different cognitive contexts. In general, functional roles are always defined in a way that is independent of the state the system happens to be in, for the idea is to capture all the possible connections between states, that is, all the possible paths through state space, not just the one the system happens to be on. This, in fact, makes it

7. To get an intuitive handle on what is going on here, it is useful to ask what *net difference* having a given representation *r* in R_Σ makes to what Σ can do, the difference it makes to the possible transitions between cognitive states. By removing *r* from R_Σ we remove all the paths through Σ's cognitive state space that involve *r*-attitudes.

8. Perhaps we can count the logical consequences of *P*, but these will be of no interest when *P* is atomic.

clear that the verificationist version of CRS isn't a version of functionalism at all: functional roles are never sensitive to the states a system happens to be in. When you write down a machine table, you have fixed all the functional roles, but you have not said anything about what state the machine is in. A functionalist version of CRS, as opposed to a verificationist version, is not vulnerable even in principle to meaning incomparability objections based on the fact that two systems, or one system at different times, typically does not have all the same attitudes.

Whether verificationist or functionalist, a coherent CRS must come to grips with the fact that it is attitudes, not representations, that have epistemic liaisons, and that epistemic liaisons can only determine representational content indirectly via the attitudes they enable.

CRS, Holism, and the Analytic/Synthetic Distinction

By a *holistic* scheme of representation, I mean a scheme in which the meaning of each representation in the scheme is dependent on the meanings of all the others. It is diagnostic of holistic schemes that adding to or subtracting from the expressive power of the scheme changes the meanings of all the representations in the scheme.[9]

If meaning is what CRS says it is, then schemes of mental representation are typically holistic: adding a new primitive evidently adds to the possible attitudes. It is plausible to assume that these new attitudes will be accessible to or from many of the old states. Evidently, every representation involved in a state that is accessible to or from the new attitude will have a new meaning.[10] Since representational schemes are holistic if CRS is true, and since CRS does not entail meaning-incomparability, it follows that holistic schemes need not involve meaning incomparability. The received view is that CRS can avoid *radical incomparability*, that is, having meaning incomparability follow from a difference in actual attitudes, only by invoking an analytic/synthetic distinction. The idea is to let only the analytic connections count toward the epistemic liaisons that determine meaning, thus insulating meaning from mere differences in "collateral information," though not, of course, from differences in analyticity-conferring

9. This definition differs from those of Fodor and Lepore (1992). Their approach is to say what it is for a *property* to be holistic, viz. that if one thing has it, then lots of things do, and then to focus on whether being meaningful, or having the content c is a holistic property. I think this approach is misleading at best. More detailed discussion of atomism and holism can be found in chapter 6. Toward the end of that chapter, I argue that since some schemes of representation are atomistic and some are holistic, meaning itself can be neither.

10. The accessibility assumption is pretty weak: a primitive representation involved in attitudes not accessible to or from any of the old states would be a pretty useless primitive: all the states involving it would be isolated from the rest of the system.

rules of inference. But we now have on the table a respectable formulation of CRS that makes meaning holistic but does not entail radical incomparability, but only a kind of moderate incomparability: systems that share a functional architecture[11] will share meanings. So what goes? Is the received view about the relation between CRS and the analytic/synthetic distinction *wrong*? Or have we smuggled an analytic/synthetic distinction into the formulation of CRS?

Doubtless the main reason for thinking that CRS requires an analytic/synthetic distinction to avoid meaning-incomparability is simply that people have in mind the verificationist version of CRS, not the functionalist version just described. The verificationist version of CRS *does* require the analytic/synthetic distinction to block meaning-incomparability, because the verificationist version does make meaning at t depend on what attitudes Σ has at t. But it is interesting to inquire whether some form of the analytic/synthetic distinction has been smuggled into the functionalist version of CRS as well.

I think it has. CRS does rely on a kind of analytic/synthetic distinction. But it is not *the* analytic/synthetic distinction. Moreover, it is an analytic/synthetic distinction that every orthodox computationalist is already committed to.

Where does an analytic/synthetic distinction come into our formulation of CRS? It comes in with the distinction between the cognitive transitions the system *can* make and the ones it actually does make. If s is on a path in state space that includes an earlier state that involves r, then Σ *can* get to s from an r-state. Of course, it may actually never do so. Whether it does will depend on whether Σ ever gets into the relevant r-state, and on whether the environment cooperates in keeping Σ on the path that leads to s. And that, in turn, depends on previous states and on the world. Still, the assumption of accessibility means that there is a possible route from the one attitude to the other, and that possibility is grounded in the fixed functional architecture of Σ. We can think of the fixed functional architecture as determining which inferences are possible in Σ and which are not.[12] The distinction between which inferences you can make given what you believe and which inferences you can make given how you are built is rather like the distinction between what is synthetically inferable and

11. Given the above conventions, we specify a functional architecture when we specify the intenders, the boxes (cognitive functions), the representational scheme (the set of available representations), and the transition function Σ on cognitive states.

12. Paths through cognitive state space are not quite inferences, of course, because they involve some nonrepresentational states, and because not every transition from one set of attitudes to another is an inference. But paths—at least reasonably short ones—through cognitive state space are inference-like; they are what inferences look like when you look at all the boundary conditions as well as the salient stuff.

what is analytically inferable. Given a representational scheme, intenders, and boxes, the processes are rather like rules of inference, dictating which attitude-to-attitude transitions are possible. From this point of view, CRS, in its functionalist guise, is just a way of saying that the fixed rules of inference—the ones that cannot be learned or unlearned because they are part of the architecture—determine meanings. On this showing, it seems fair enough to say that a functionalist formulation of CRS does rely on a kind of analytic/synthetic distinction.

Those, including me, who don't like the analytic/synthetic distinction, are fond of saying that you can't tell rules from premises. The fact that one goes inevitably from 'bachelor' to 'unmarried' could be due to a rule of inference, making the connection analytic, but could just as well be due to a readily available premise to the effect that all bachelors are unmarried, making the connection synthetic. How is one to know? But in the case of a well-specified computational system, one *does* know, for one knows the difference between fixed functional architecture—the processes, intenders, boxes, and representational scheme that define the system—and the attitudes the system happens to have at a given time—which applications happen to be in which boxes. Anyone who wants to distinguish in this way the structure of the mind from its contents is committed to enough of an analytic/synthetic distinction to save CRS in its functionalist form from radical meaning incomparability. Of course, not everyone is committed to a principled distinction between architecture and content, process, and data. Connectionists, for example, are not. I am not sure how many will be willing to embrace connectionism in order to be in a position to defend the standard objection to CRS. Those who do like to keep a clear distinction between what we believe and how we are built will have to find some other way to bash CRS, or they will have to save their criticism for the verificationist version.

The meaning-incomparability objection to CRS, as it is usually formulated, is seriously confused. Meaning incomparability does not follow from a change in attitudes. But it does follow from a change in representational primitives. Adding (or subtracting) a primitive from a system's representational scheme will alter the meaning of almost everything if it makes any difference at all to the system's cognitive capacities.[13] It will follow that, strictly speaking, new primitives cannot be learned by any inferential process, since there will be no coherent way of describing the inference. Holistic representational schemes generate this kind of incomparability by definition, and CRS makes every scheme holistic. But it is not clear how bad this kind of incomparability really is. Fodor, who does not like holism (Fodor and Lepore, 1992), is notoriously on record

13. The same goes for adding (or subtracting) a new (elementary) process, intender, or box.

in support of the view that primitives cannot be inferentially learned anyway (Fodor, 1975). It is true that much of our practice in theory and model building in computational psychology assumes an atomistic scheme of representation, in that we regularly add or subtract primitive representations and processes without supposing that everything changes meaning when we do so. But perhaps this is just bad practice. Philosophy, at any rate, cannot simply assume atomism on the grounds that standard practice seems to presuppose it. Philosophy is, I am sometimes told, supposed to earn its keep by being critical of the unexamined presuppositions of standard practice. We are going to have to forget the incomparability objection to CRS and look elsewhere if we are to see what is wrong with use theories of meaning in general, and CRS in particular.

How CRS Makes Content Explanations Trivial

Here is the plot for the rest of this chapter. I argue first that CRS trivializes explanatory appeals to content. I don't really expect advocates of CRS to disagree with this, but I think it is useful to have the consequence explicitly drawn. I then show that CRS cannot accommodate the distinction between target and content, and hence cannot support a robust notion of representational error. At bottom, it is this failure to come to grips with error that unhinges content explanations for CRS. I also show that CRS cannot distinguish between representation and detection, and relate this to the deflationary approach to content implicit in CRS. I conclude by arguing briefly that explanatory appeals to content are not trivial, as CRS implies they are.

In the story rehearsed in chapter 2 featuring the chess system Σ, we seem to suppose that R_{P3} has the consequences it does in part because of what it represents. Yet the consequences of tokening R_{P3} in these circumstances are part of the epistemic liaisons determined by R_{P3}, hence, according to CRS, part of the specification of R_{P3}'s representational content. It follows that CRSers cannot explain

(D) Σ conceded a draw to black,

by appeal to,

(E) When the position was P1, Σ erroneously believed that the position after M would be P3,

because CRS renders this explanation trivial. To see this, note that what (E) says, according to CRS, is this:

(E1) Σ's applying R_{P3} to the position-after-m in the belief box yields an attitude A with the epistemic liaisons $\Sigma(A)$, that is, the set of epistemic liaisons associated with A in Σ.

Now (E1) simply *entails* that Σ concedes a draw to black, for conceding a draw in this situation is one of the epistemic liaisons definitive of A. So (E1) entails (E2):

> (E2) The occurrence of A in the situation in question has the consequence that Σ concedes a draw to black,

which is evidently empty as an explanation of (D).

This should come as no surprise. According to CRS, talk of an attitude's content is just shorthand for, among other things, talk of the consequences of its occurrence. So, of course, CRS won't allow you to explain the consequences of the occurrence of an attitude by appeal to that attitude's content. CRS thinks about content as we might think about valence. Imagine a theory that tells us what bonds with what in what proportions. We could simply list all the possibilities (assuming they are finite). Or we could do this: assign a positive or negative number to each radical, and state the following rule: any combination that adds up to zero is a compound.[14] Valences are a kind of fiction in this theory (multiply them all by a constant, and the theory remains unchanged): Specifying something's valence is simply a convenient way of specifying its bonding potential without actually having to mention all the other elements and all the proportions explicitly. Given this fact, it follows that you cannot explain why oxygen bonds with hydrogen in the ratio of one to two by appeal to the valences of oxygen and hydrogen.

CRS gives us a valence theory of content. Content, on this view, is really a kind of fiction, for it treats talk of content as just a convenient way of referring to the epistemic liaisons enabled by a representation without having to mention them explicitly (a useful device, since we typically don't know what the relevant epistemic liaisons are). Once we realize this, it is obvious that CRS isn't going to allow us to explain behavior by appeal to content any more than valence theory is going to allow us to explain bonds by appeal to valence. In both cases, what we wind up with is a kind of thinly disguised joke à la mode de Molière rather than a genuine explanation. CRS trivializes content explanations.[15]

14. This won't actually work, but it might have. Pretend it does work for the sake of the analogy.

15. Block (1986, pp. 668–669) offers two reasons for thinking that CRS doesn't trivialize content explanations: (1) We are talking about a complex disposition that relates the attitude in question to a host of other consequences (and antecedents) besides conceding a draw to black; (2) what Armstrong (1968) called the categorical basis of the disposition can be identified independently of the consequence in question. Neither is persuasive.

1. You can't explain why opium puts people to sleep by appeal to its dormitival virtue. This situation is not mitigated by noticing that opium also makes people high. If making the disposition complex really changed things, valences would explain bonds!

A lot of you won't be bothered by this, because a lot of you think content is *causally inert* anyway. According to you, content explanations never were in the cards, so their being undermined by CRS is a virtue of CRS, not a vice.

A natural first response to content inertness is this: You can't sustain the position that content is inert when you leave the abstracted heights of philosophy and look at what is actually happening. When I tell you that LOOK-AHEAD comes up with R_{P3}—when I actually show you R_{P3}—you can see that we have got a case of error here, because you can see that R_{P3} represents P3, whereas it ought to represent the position after *M*, viz. P2. I've set this up for you so that it is easy to see. You can't miss it. And, of course, there's no getting around the fact that it is this error, that is, this misrepresentation of P2 as P3, that causes the trouble.

Though I am sympathetic with this line of response to content inertness, I don't think defenders of CRS need be terribly impressed. They can reply as follows:

> We can see what R_{P3} represents *to us*; but we cannot see what R_{P3} represents *to* Σ. For all we know, just looking at R_{P3}, it *does* represent P2 to Σ. The only thing that tells us that we and Σ understand the same thing by R_{P3} is that Σ does such things as mistakenly concede a draw to black in the situation we've been discussing. That is, the only thing that gives us a line on what R_{P3} represents to Σ is the set of epistemic liaisons R_{P3} determines in Σ, and that is just what CRS predicts. The idea that content specifications do more than simply sum up epistemic liaisons as valence specifications sum up chemical liaisons is just an illusion, an illusion created by the fact that the representations in our example have meanings to us that are independent of the epistemic liaisons those representations enjoy in Σ.

Given the availability of this reply, I don't think a brute appeal to the intuition that content explains behavior is going to move the confirmed advocate of CRS to skepticism about the content inertness CRS entails. We have to do better.

I don't want to argue about the causal status of content. But I do want to argue for its explanatory relevance.[16] If CRS is right, there is no point

2. Arguing that the categorical basis of a dispositional property explains the disposition's manifestations amounts to conceding that it is not the dispositional property that explains manifestations. You cannot wriggle out of this by identifying a dispositional property with its categorical basis when, as in the case at hand, the disposition in question can be multiply realized—that is, has different categorical bases in different systems or in the same system at different times.

16. All explanation is not causal explanation. See Cummins (1983) for a sustained attempt to show this.

beyond descriptive convenience in the semantic interpretation of the states of a cognitive system. A genuine representationalist, on the other hand, thinks that the status of certain states as representations is of real theoretical importance. If we can show that representationalists are right about this, that is, show that representational content, and hence attitude content, has a genuine explanatory role, then we will have refuted CRS.

CRS and Representational Error

We have shown that CRS trivializes content explanations, but we have yet to determine whether this is a virtue or a vice. My approach to this issue is a bit indirect. Before we get to the explanatory relevance of content, I want to have a look at how CRS deals with error and the distinction between representation and detection. It is the notion of representational error—or, rather, the distinction between error and correctness—that gives content explanation its bite. Intuitively, as I said recently, it is the fact that R_{P3} does not represent what it is supposed to that explains Σ's flawed performance in our running chess system example. I'm going to try to make this intuition do some work by arguing that CRS cannot support a robust conception of error. The basic idea is that CRS will always force you to redescribe an alleged case of error as a case of misinterpretation: any evidence of representational error is better evidence for a different assignment of contents.

The idea that Σ made an error and the idea that Σ is a chess system are based on the same thing, viz. a certain interpretation of Σ's representations, targets, and boxes. But we cannot take that interpretation for granted if, as my imagined CRSer lately claimed, epistemic liaisons provide the only line on what these things mean *in Σ*. Perhaps we should say that Σ is a schess system, and that what it did is a winning move in schess, not a losing move in chess.

Must CRSers take this seriously? I think they must, but it isn't obvious. It isn't obvious because, as remarked earlier, CRS assigns a content to an attitude by assigning a set of epistemic liaisons to it, and this leaves us with no specification of how that content should be expressed. A fully articulated CRS "interpretation" of Σ would leave us entirely in the dark as to whether R_{P3} represents P3, because it would leave us totally in the dark as to whether the attitude that occurred when Σ tokened R_{P3} on the occasion in question had the content *that the position after M is P3*. We can put this point by saying that CRS assigns content determinations to attitudes without actually interpreting them. And in the absence of any interpretations, who is to say whether Σ plays chess or schess, or whether the attitude of interest involved representational error or not?

Surprisingly, we can answer this last question: *There cannot be any error according to CRS.* Think of all the epistemic liaisons determined by R_{P3}: They determine the "use" of R_{P3} in Σ, that is, they determine what targets R_{P3} will be applied to by Σ. Now there must be some factor that fixes interpretation that is independent of use, otherwise there is no sense in asking whether Σ uses R_{P3} erroneously or correctly. We must, in short, have the distinction between what Σ uses *r* to represent on a given occasion (target), and what *r* actually represents (content) if we are to make sense of the idea that Σ can misuse *r*. We have to have a distinction between representational content and target to have error. CRS defines the content of R_{P3} in Σ as its use by Σ, so there can be no question as to whether Σ uses R_{P3} erroneously or correctly.[17] Since, according to CRS, the actual use of R_{P3} in Σ defines the content of R_{P3} in Σ, Σ must use R_{P3} correctly.

So far as I can see, there are only two ways to try to get around this problem: Idealization, and Relativization. Let's look at each in turn.

Idealization

A CRSer might try to allow for error by supposing that content is fixed by ideal use rather than actual use. The trouble with this idea is that it is circular. You cannot understand ideal use as correct use because, in the absence of some criterion of error, you don't know what to idealize away *from*. Σ may be an ideal schess player as it stands. Nor can you identify ideal use with use leading to success or ideal performance. You don't know whether you have success until you know whether you have a chess system or a schess system, and you don't know *that* until you have fixed an interpretation. Put another way, you cannot rule that the bad inferences don't count toward content determination because you don't know what the bad inferences are until you have fixed an interpretation.[18]

A number of moves have been made in response to this difficulty.

Ideal Use Is Rational Use A number of writers (Dennett, 1987; Davidson, 1973; Pollock, 1989) have tried to salvage the idea that correct use is ideal use by arguing that ideal use can be independently identified as the use required by rationality. The argument goes this way: Intentional explanation presupposes rationality. Since failures of rationality undermine

17. Loui (1991) makes a related point when he argues from what amounts to an implicit assumption of CRS to the conclusion that a system that infers '$(x)Gx$' from 'Ga' and the absence of known counterinstances isn't engaged in ampliative inference. The argument is that, in such a system, writing 'Ga & $-(Ex)$ (it is known that $-Gx$)' is just another way of writing '$(x)Gx$'.

18. You do not want to identify correct inferences with those leading to success in any case, because, as we shall see shortly, misrepresentations and incorrect inferences may, in certain circumstances, be more "successful" than correct ones.

intentional explanation, interpretation is pointless in the absence of rationality. Rationality therefore constrains interpretation. We don't have to worry that we have a perfect "schreasoner" on our hands rather than in imperfect reasoner, because schreasoners do not have intentional states at all. So long, then, as we cleave to intentional explanation, we can take failures of rationality as clues to representational errors.

There are three problems with this line of thought. The first is that rationality cannot be specified in a way that is independent of the facts about how a cognitive system actually works, including the facts about its representational capacities. Rationality cannot, for example:

- Require beliefs that are impossible given a system's representational capacities
- Require beliefs that would take too many resources to acquire or justify
- Require beliefs that would be computationally intractable to use

But if what is rational cannot be abstracted from what is psychologically possible, then rationality cannot provide an independently fixed point that can be used to leverage interpretations. What is rational for a system depends on that system's cognitive architecture, and that, in turn depends on its representational capacities. The proposal under consideration requires, impossibly, a conception of rationality that is logically prior to what a system can or does represent.

The second problem with the idea that representational errors can be identified via deviations from ideal rationality is that content explanations are not, in point of fact, limited to, or even primarily aimed at, the sorts of things that are appropriately assessed for rationality. Indeed, as Fodor has emphasized in another connection (1983), representational explanation has been most successful in the case of processing modules of the sort generally supposed to be involved in early vision and in language processing. These modules are precisely not general-purpose reasoners, and are certainly not straightforwardly subject to epistemological assessments of the sort that are supposed to anchor interpretation.

The third problem with the idea that representational errors can be identified via deviations from ideal rationality is that the explanatory interest of representation is undermined if you define it in terms of rationality. Representationalists want to explain the capacity for cognition generally, and for rationality in particular, by appeal to a prior capacity for representation. This move is undermined if representation is in turn explained in terms of rationality. This, as we noted earlier, doesn't worry the CRSer, because CRSers are antecedently convinced of the explanatory idleness of representation. But it is important to see that the trivialization of representational explanation and the CRSers' typical treatment of error go hand in hand. CRSers aren't trying for a *serious* treatment of error, because they aren't trying for a *serious* treatment of representational explanation in the first place. They are only trying to avoid the embarrassment of having to admit that there isn't any such thing as representational error. But, as Perlman points out (1993), they might as well admit it: the only reason to take representational error seriously is because you take representational explanation seriously. If you don't do the latter, you needn't bother with the former.

Competence and Performance Distinguishing ideal from actual use is familiar in linguistics as the distinction between competence and performance emphasized by Chomsky (1959). The idea, popularized by Fodor (1975), that mental representations are expressions in a language of thought, encourages a similar distinction between competence and performance in the use of mental representations. It is instructive, therefore, to see why a competence/performance distinction is unavailable to the theory of mental representation.

Linguistic competence is simply ideal linguistic performance, that is, the performance that the system would exhibit but for resource limitations, physical breakdown, and interference from other processes. In computational terms, the underlying assumption is that we are dealing with a system that implements a perfectly general and correct algorithm for language use, but which has limited resources and is dependent on the reliability of its physical instantiation. We might conceive of ideal use of mental representations, then, as the use that would occur but for physical breakdown or resource limitations (assuming these can be identified independently of content assignments). But this is of no help in the present context unless error can be identified with departure from ideal use in this special sense. It is clear, however, that representational error is not always, or even typically, due to breakdown or limited resources. It is much more commonly due to the execution of perfectly good inferential routines that happen to come up with the wrong result because of incomplete or misleading information. These are the uses that are justified but incorrect. If you insist that every use that is unaffected by resource limitations and breakdown is correct, you will be unable to even articulate the distinction between justified use and correct use. You will, in short, become a verificationist. Whatever one thinks of verificationism generally, one doesn't want to be a verificationist about mental meaning, for one wants to be able to take seriously the possibility that cognitive systems are not optimally designed, let alone endowed with a design that is foolproof but for resource limitations and breakdown. CRS will force us to see an ill-designed chess system as a well-designed schess system.

Correct Use Is Adaptive Use A variation on idealization is to adapt an idea of Millikan's (1984) and construe error as the gap between actual use and Normal use, where Normal use is understood as whatever use was historically crucial in securing the replication of the mechanisms that produce or consume the representations in question, or do both. Correct use, on this conception is, as it were, adaptive use. This approach is attractive because it suggests that a system uses a representation correctly when it uses it in whatever way it *needs* to use it to succeed in the game of life. It seems plausible to suppose that it was those occasions on which representations of edges were applied to edges that made edge representations useful. Hence, it seems plausible to suppose that *r* represents edges just in case its useful applications were to edges.

Note the two-step strategy here. First, representational content is defined in terms of *correctness*, the idea being to identify *r*'s content with its target on occasions of correct use: *r* represents edges just in case it is correct when its target is an edge. Second, correctness is defined in terms of adaptation: correct uses are adaptive uses. This reverses the strategy I have been pursuing, which defines correctness in terms of content and target, and hopes for independent definitions of these. Either way, however, error turns out to be a mismatch between target and content, and both strategies require an independent treatment of target fixation.

There is no objecting to the truth of step 1: it is tautological that *r* represents edges iff it is correctly applied to edges. The issue is rather one of explanatory strategy, since step 1 is of no use unless one can define correctness independently of content. The burden thus falls squarely on step 2, the step that explains correctness as adaptive use. There are three reasons why I don't think this will work.

1. *Adaptiveness does not correspond to correctness.* There are two sides to this: *A* representation can be correct but not be adaptive, and it can be adaptive without being correct. (a) *Correct but not adaptive.* You can design a trout that, in spite of refraction, correctly represents the positions of insects flying just above the water, but this will not be adaptive in a trout already equipped with a jumping routine that compensates for refraction. (b) *Adaptive but not correct.* These are the cases in which the errors are not *serious.* A system that consistently represents little ambient black particles as insects will serve a trout well in an environment in which enough of the little ambient black particles are insects, and few are harmful. It is cheaper to design a tolerant digestive system in such circumstances than it is to design an insect recognition system free of false positives. Efficiency often entails inaccuracies in a resource-bounded system. In such systems, inaccuracies that are efficient but not serious will be more adaptive than more accurate but less efficient alternatives.

2. *Adaptational stories cannot distinguish correlated contents.* This is the point Fodor makes against teleological theories (Fodor, 1990b). If being F and being G are correlated in an organism's environment, the adaptiveness of representing F and of representing G will be the same (provided the computational costs are comparable). Fodor points out that evolutionary theory won't distinguish between a story in which a representation r in a frog's visual system represents flies and one in which it represents ambient particles provided that r triggers the snapping mechanism, and provided that there is enough of a correlation between ambient particles and flies. The correlation need not even be lawlike; it can be pure coincidence. Coincidences will get selection jobs done just as well as laws, provided they last long enough, as every operant conditioning experiment shows.

3. *Adaptational stories get the explanatory order wrong.* Representations, when they are adaptive, are adaptive because they represent what they do. Think of cognitive maps (Tolman, 1948): They are adaptive because they are isomorphic to the spaces they map; they are not isomorphic to the spaces they map because they are adaptive.[19] Explaining correctness in terms of adaptation gets matters backward. The adaptational theory is motivated by the fact that it is plausible (though not quite right; see [1]) to suppose that correct uses of a representation are adaptive. But what makes this plausible is the idea that the uses in question are adaptive because they are correct. That motivation is undermined on the assumption that the uses in question are correct because they are adaptive. To a first approximation, representationalists want to explain cognitive success in terms of representational correctness. But you cannot do *that* if you explain representational correctness in terms of the adaptiveness of cognitive success.

The idea that the correct use of a mental representation is its ideal or Normal use is a dead end, then. The alternative is to relativize. If you can't find a principled distinction between saying a system is playing chess and making representational errors and saying it is playing schess and getting everything right, then make a virtue of necessity: say it is both. It is chess relative to one interpretation (I_c) and schess relative to another (I_s).[20] When Σ applies R_{P3} to the position-after-m, it makes an error relative to I_c, but not relative to I_s. This allows for error at the price of abandoning CRS altogether. For this suggestion amounts to saying that, if we fix con-

19. You cannot get around this by arguing that what makes the isomorphism representational is the fact that it has a selection history. Since it is the isomorphism itself that does all the explanatory work, insisting that it isn't a *representation* of the space it maps until it has a selection history just prices you out of the market, for it makes the property of being a representation explanatorially irrelevant. For an extended treatment of this theme, see chapter 7.
20. This, I blush to admit, was the route I took in *Meaning and Mental Representation*.

tent assignments in advance somehow, then we can tell error from correctness, and hence can tell success from failure. While this is true enough, it is clear that epistemic liaisons are no longer playing a role in determining content.

The Nontriviality of Representational Explanation

CRS offers us a valence theory of meaning which more or less self-consciously trivializes representational explanation. This, of course, is an embarrassment to CRS only if representational explanation is not trivial.

The difficulty is not, of course, that CRS gets the causal properties of representation wrong. Indeed, CRS gets them right by definition since it, in effect, identifies *r*'s content with its relevant causal properties. As long, then, as we think that what we are after is an accurate picture of the causal structure, it will be impossible to see how CRS could possibly miss something of explanatory importance.[21] It follows from this observation that the explanatory role of content is not to be found in its causal role. To see where it is found, you have to appreciate that it is the notion of representational error that gives representational explanation its bite. Representational error, as we have seen, depends on a distinction between how a representation is used, and what it means. Since use theories explicitly deny this distinction, they undermine the notion of representational error and with it the explanatory importance of representation.

As we saw in chapter 3, it is the distinction between representational correctness and error, or, better, between representational accuracy and inaccuracy, that allows one to exploit the distinction between the degree of inaccuracy and its seriousness, and correlatively between effective representation and accurate representation. And it is what allows one to exploit the distinction between forced and unforced error, and between forced error and the inability to exploit an adequate representation or scheme. Let's look briefly at what this is going to cost.

Effectiveness vs. Accuracy

As we saw in chapter 3, less accurate representations can sometimes be more effective than more accurate alternatives because of the computational costs in achieving accuracy in the first place, and in manipulating highly accurate representations once they are achieved. Accurate predator recognition is bound to be slow, and, given the relatively insignificant

21. This, I think, is the key to understanding Stich's syntactic theory of mind (1983). If you are an orthodox computationalist, then you think that the causal properties of representations turn on their syntax. From this, and the assumption that it is causal structure we are after, it simply follows that content is irrelevant.

consequences of false positives, a faster system that generates many false positives is likely preferable. Without a distinction between a representation's use and its meaning, you cannot coherently articulate this simple observation. Any reason to suspect false positives will be a better reason to suspect misinterpretation, for applying a representation to a passing cloud or the shadow it makes will count as a misuse only against some independent standard of what the representation means and hence of how it should be used. In this context, it is easy to see why adaptationist standards of correct use cannot be right. For the use of *r* that accounts for the replication of its producers/consumers need not be the right use on any particular occasion. Indeed, given the efficiency of a policy that tolerates false positives, applications to nonpredators surely were among the uses that contributed critically to the replication of the users and consumers of that representation. *The point about the efficiency of a policy that tolerates false positives is that it is adaptive.* It follows that appeals to adaptiveness cannot distinguish correct uses from incorrect uses. Use theories undermine the important distinction between accuracy and effectiveness because that distinction requires a notion of representational content that is independent of use.

Forced and Unforced Error

We arrive at a similar conclusion when we reflect on the distinction between forced and unforced error. To take an example familiar from the history of philosophy, the Kantian position that the human visual system constructs Euclidean images makes the introspectively plausible prediction that we will be unable to imagine or see non-Euclidean spaces accurately. The Kantian position treats the visual representation of non-Euclidean space as a case of forced (but not serious) error.[22] If we adopt a use theory of meaning, however, we will be unable to articulate this position and the issues it raises, for we will have to suppose that visual representations, since they are never[23] applied to non-Euclidean space, are themselves non-Euclidean in content! Use theories cannot contemplate the possibility of forced error because that requires representations that never correctly apply.

In just the same way, use theories will be unable to construe the dispute between Piagetians and their opponents concerning the shifting patterns of mistakes developmental psychologists observe in the performance of children on word arithmetic problems, for that dispute just is a dispute about whether the mistakes are due to forced representational error or to

22. Kant, of course, didn't think Euclidean representation of space was erroneous, but we do.
23. Or almost never—certainly never in detection cases (causal theories) or cases with selective significance (adaptationist theories).

developmentally significant changes in the capacity to exploit adequate innate representational sources (see D. Cummins, 1988, 1991). Use theories are bound to see this as a pseudo-issue founded on misinterpretation. Since they are bound to collapse the distinction between forced error and the inability to exploit existing representational resources, they will see the relevant development as either a case of meaning change (verificationist version) or as a kind of frequency effect in which the frequency of certain uses drops and the frequency of others rises as a function of cognitive changes (functionalist version). The hypothesis, shared by most parties to the dispute in the developmental literature, that systematic mistakes are due to systematic misrepresentation, is simply not available to use theories.[24]

The Moral of the Story

Suppose I'm right: cognitive science requires a robust notion of representational error, and use theories cannot provide one. *Why* do we need the notion of representational error? It is one thing to see that certain scientific issues presuppose the notion of representational error, as we've just done. That shows *that* we need misrepresentation, but not *why*. It is quite another thing to have a philosophical account of the role of representation in explanation that allows us to see why a scientific understanding of cognition requires a serious conception of content, one that makes misrepresentation possible by keeping meaning and use apart.

To see what representationalism really buys you, you need a theory of representation. I'm going to try to provide one in chapter 7. But even without a theory of representation in hand, you can see enough to keep you moving in the right direction. Appeals to representation buy two different things.

Dimension Shifts Semantic interpretation effects a dimension shift that reveals mere computation as something epistemically assessable: the manipulation of formally individuated objects is revealed as addition, or planning, or language processing. Since cognitive capacities are capacities for epistemic constraint satisfaction, and since epistemic constraints are defined semantically, that is, defined over semantically individuated objects, explaining cognition by appeal to computation requires semantically interpreting the objects computed.

This much is available to the advocate of CRS. This is the account of the explanatory role of content I offered in *Meaning and Mental Representation* in connection with the functional role theory of content offered

24. Carey (1985) clearly endorses some form of CRS, yet helps herself to the notion of representational error. As she must: developmental psychology without representational error is hopeless.

there (Cummins, 1989). If the sort of dimension shift just described is all one sees of the explanatory role of representation, then CRS can seem a very plausible story. Imagine a device, m, that looks like a calculator. When m is given '19', '18' and '25' as inputs, *m* generates a '2' in the rightmost position of the output buffer. At this point, *m*'s target is the current carry, 20. Device *m* constructs '10', and generates a '5' in the next output position. Given these facts, we could say the thing is an adder that has misrepresented a carry, or we could say that it is a shmadder (like an adder except...). There appears to be little to choose between these stories. There is no *substantive* dispute here about what *m* does, and we can explain the performance of an embedding system in terms of the incorporation of a faulty adder, or in terms of the incorporation of a successful shmadder, for, in both cases, the point is to contrast the performance we actually get with the performance we would get if we incorporated a successful adder.

Reasoning vs. Representation As long as we focus on cases like this in which semantic interpretation enters only to bridge the gap between, say, a description in terms of symbols and one in terms of numbers, the notion of representational error plays no significant role, and so use theories seem perfectly adequate. What, then, does the notion of misrepresentation buy us? Or equivalently, what do we gain from a clear distinction between what a representation means (content) and how it is used (target)?

A use-independent notion of content or accuracy allows one to distinguish the correctness of a representation from the correctness of reasoning (or other processing) done with it. If we misrepresent the facts to Holmes, he will misidentify the culprit, even though he reasons flawlessly. Similarly, distorted memories about parents may mislead an excellent planner, resulting in botched plans for a wedding. CRS, however, must see this as inadequate *planning*. If we follow CRS in this, however, we will have a hard time explaining why plans that don't involve taking into account the personal characteristics of family members work fine. Distinguishing meaning and use allows us to factor cognition into representation problems and reasoning problems. The flexibility that results from allowing for the interaction of these independent factors is what gives representationalist accounts of cognition the descriptive and explanatory power needed to understand the value of fast but error-prone predator recognition systems, or to understand how to reconcile developmental stages in the ability to solve word arithmetic problems with the impressive evidence for the innateness of arithmetical concepts (Wynn, 1992).

Conclusion

The standard objection to CRS is off target. CRS does not, when given a functionalist rather than a verificationist formulation, have the consequence that systems with different beliefs are meaning-incomparable. The consequence of radical meaning-incomparability is blocked by the mediation of a kind of analytic/synthetic distinction, but it is a kind of analytic/synthetic distinction that every orthodox computationalist must accept, for it is just the distinction between what Σ can do given its architecture, and what it can do given its current attitudes. CRS *does* have the consequence that meaning is holistic—that systems with different primitives are incomparable—but it isn't clear whether that is a bad consequence. It would mean that we could not describe acquisition of a new primitive as inferential learning, but Fodor (1975) has argued persuasively for that on independent grounds anyway.

CRS is wrong, but not because it engenders radical meaning-incomparability. It is wrong because it trivializes content explanation. It trivializes content explanation because, like all use theories of meaning, it cannot support a robust notion of representational error, and its explanatory potential is thereby deprived of a crucial degree of freedom. Use theories have trouble with representational error because the concept of representational error, and hence of the misuse of a representation, requires a distinction between how a representation is used and what it means. Use theories all have a two-step structure: They identify r's content with its target in cases of correct use, and then they attempt to specify some naturalistic sufficient condition of correct use. It is the second step that causes the trouble, and it is not hard to see why. Under what condition is the use of a representation guaranteed to be correct? The history of epistemology should teach us that nothing *guarantees* accuracy. Or rather: The only conditions that guarantee accuracy also trivialize content. For example: You can make believing a sense datum statement a sufficient condition for its truth, but the price you pay is that you empty it of evidential value. If s is evidence of something else, t, then any evidence against t is, indirectly, evidence against s. If s is to be immune in principle to counterevidence, it cannot be evidence for anything else. Epistemological conditions of correct use (e.g., r is correctly applied to t by Σ if Σ knows that t satisfies r) are bound to violate the *Explanatory Constraint* laid down in chapter 1, since we want representation to explain cognition and not the other way around. Nonepistemological conditions of correct use are bound to fail, for nature never guarantees accurate representation.