Representations, Targets, and Attitudes

Robert Cummins

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Palatino by Asco Trade Typesetting Ltd., Hong Kong and was printed and bound in the United States of America. This book was printed on recycled paper.

Library of Congress Cataloging-in-Publication Data

Cummins, Robert.
Representations, targets, and attitudes / Robert Cummins.
p. cm.—(Representation and mind)
"A Bradford book."
Includes bibliographical references and index.
ISBN 0-262-03235-X (hc : alk. paper)
1. Representation (Philosophy) 2. Mental representation.
I. Title. II. Series.
B105.R4C84 1996
128'.2—dc20

95-22502
CIP

Contents

Chapter 1
Introduction

Contemporary theories of cognition make heavy use of the concept of mental representations. These theories all share the view that cognizing is processing mental representations. As a result, *the* topic in the philosophy of mind for some time now has been to explain the relevant notion of mental representation—to say, in some illuminating way, what it is for something in the mind to represent something. We can put this problem in perspective by comparing it to other problems that arise in connection with mental representation.

The Problem of Mental Representation

The difficulties surrounding mental representation can be ranged under four heads:

1. *Content:* Which contents are represented in the mind? What information is actually represented in the mind when it goes about its cognitive business?
2. *Form:* What form does mental representation take? There is the very general issue of what kind or kinds or representational scheme or schemes the mind employs (e.g., images, symbolic structures, activation vectors), and there are many particular issues concerning which of the available forms within a given scheme are employed in a given task.
3. *Implementation:* How are the mind's representational schemes implemented in the brain?
4. *Definition:* What is it for one thing to represent another?

The first three problems are, in various degrees, problems for empirical research. The fourth is a philosophical problem. But, in spite of this (rather artificial) division between empirical science and philosophy, answers to any one of the four questions constrains answers to the others. Answers to (1) and (2) are mutually constraining because a given content may be representable in one scheme but not in another, or because a given

content may be represented efficiently in one scheme, and not in another. Answers to (2) and (3) are mutually constraining because some representational schemes may not have neural implementations, or may not have neural representations in our brains. And the first three issues evidently constrain the issue of definition, for a good definition of representation should be compatible with the contents, forms, and implementations that mental representations are known to take.

Since our problem in this book is going to be the problem of definition, it is worth taking explicit notice of the way in which (1) through (3) constrain (4).

As I see it, the most fundamental substantive constraints on a philosophical theory of mental representation (i.e, on a definition) are these:

- *Explanatory Constraint:* The theory should underwrite the explanatory appeals that cognitive theory makes to mental representation.
- *Implementation Constraint:* The theory should be compatible with the best scientific stories about what sorts of things actually do the representing in the mind/brain.

There are several representationalist theories of cognition currently on offer, of course. The possibility is very much open that different theories give rise to different explanatory and realization constraints. It is supposed by many, for example, that connectionist architectures place special constraints on both the kinds of representational schemes that are possible (no language of thought), and on the explanatory role played by representation (computation, being local, is not defined over the representations, which are distributed) (Smolensky, 1987; Fodor and Pylyshyn, 1988; Fodor and McLaughlin, 1990). A single, unified theory of mental representation might, therefore, be impossible. Perhaps science will settle on one approach to cognition eventually, but this seems unlikely to happen in the near future. Meanwhile, we shall just have to limp along with what we have, keeping in mind that our efforts may all be obsolete one day, just like the science on which they are parasitic.

Notice that among my fundamental constraints on a theory of representation, one does *not* find a constraint to the effect that a theory of mental representation should explain ordinary content attributions, such as "Peter believes the Normans invaded England in 1066." It would be nice to know what beliefs and desires are, and what it is to believe or desire that p rather than q. I will, in fact, have something to say about this. But the problem of mental representation, as I see it, is not primarily the problem of explaining the kind of content attributions that figure in ordinary attitude talk. The reasons for this will emerge as we go along, but it is well to be forewarned.

Naturalizing Content

A good deal of the recent philosophical literature on mental representation takes the problem to be one of naturalizing content. The motivation here is ontological: one wants representation defined in a way that makes it clear that the representation relation is, unlike demonic possession or spell-casting, a naturalistic relation of the sort countenanced by contemporary science. In practice, the operative constraint is that representation should be defined in a way that avoids the use of semantic terms such as 'means', 'refers', 'true', and also avoids intentional terms such as 'believes', 'desires', and 'intends'.

My ontological conscience is pretty weak. I'm no substance dualist, but I find propositions, possible worlds, numbers, and sets unavoidable. Since I have no idea how to be a hard-headed ontological physicalist about these, I feel uncomfortable about banging the table over the ontological status of representation. However, I do think that the standard practice of insisting on a definition in nonsemantic and nonintentional terms is quite proper on methodological grounds. A definition of representation in semantic terms would simply be circular. The term 'representation' is just a generic semantic term. If we don't avoid other semantic terms in its definition, we have no very good reason to suppose the definition makes any worthwhile intellectual progress.

The stricture against intentional terms requires a bit more comment. I think the general idea is that "*t* is a thought *that p*" involves a relation between *t* and the proposition *that p* that is mysterious in just the same way as the relation between a representation and what it represents: a meaning relation of some sort. On many accounts (e.g., Fodor, 1975, 1987), thinking *that p* requires representing the proposition *that p*. On almost every account, including my own, thinking *that p* requires representing something, though it may not require representing the proposition *that p*. If any of these accounts of thought are on the right track, defining representation in terms of thought would be regressive at best, and that is enough to motivate the stricture against intentional terms in the definition of representation. But there is a more general reason as well. Current cognitive science typically seeks to explain cognition by appeal to representation. This explanatory strategy is undermined if representation cannot be defined in a way that makes it independent of cognition. For example, a definition of representation in terms of rationality or inference (Dennett, 1987; Pollock, 1989) evidently threatens to undermine the standard explanatory strategy. By the same token, the capacity to think, believe, intend, plan, and desire is a cognitive achievement that is, or presupposes, the very sorts of phenomena that appeals to the processing of representation are supposed to explain. It is commonly held that only (at least minimally) rational creatures can be said to have beliefs or plans

(e.g., Pollock, 1989). We do not want to explain belief in terms of rationality, rationality in terms of representation, and representation in terms of belief.

The upshot is that the widespread idea that representation should be defined nonsemantically and nonintentionally is a corollary of what I called the *explanatory constraint* above. *Representation* is an explanatory primitive in a science that seeks to explain mental phenomena generally, and cognitive phenomena in particular. We do the foundations of this science no service, then, if we define representation in mental or cognitive terms.

A final word needs to be said about prerequisites: I m going to assume familiarity with the recent philosophical literature on mental representation. This is not intended as a text. In the same spirit, though with less confidence, I am going to assume passing familiarity with contemporary cognitive science. This seems fair enough: philosophers of physics assume familiarity with physics, so philosophers of cognitive science ought to be able to make the analogous assumption. There is a regrettably widespread assumption in philosophy that one can be a philosopher of mind without knowing what science has to tell us about the mind. We seem to have got over that sort of a prioristic isolationism in the rest of the philosophy of science (though I suspect some philosophy of language still goes on in splendid isolation from linguistics); it s time we got over it in the philosophy of mind. The idea that the philosophy of mind generally, and the theory of content in particular, is a branch of the philosophy of science, is not exactly universal. Perhaps it is contentious. So be it. I just can't imagine what else a respectable philosophy of mind could *be*.

(e.g., Pollock, 1989). We do not want to explain belief in terms of rationality, rationality in terms of representation, and representation in terms of belief.

The upshot is that the widespread idea that representation should be defined nonsemantically and nonintentionally is a corollary of what I called the *explanatory constraint* above. *Representation* is an explanatory primitive in a science that seeks to explain mental phenomena generally, and cognitive phenomena in particular. We do the foundations of this science no service, then, if we define representation in mental or cognitive terms.

A final word needs to be said about prerequisites: I m going to assume familiarity with the recent philosophical literature on mental representation. This is not intended as a text. In the same spirit, though with less confidence, I am going to assume passing familiarity with contemporary cognitive science. This seems fair enough: philosophers of physics assume familiarity with physics, so philosophers of cognitive science ought to be able to make the analogous assumption. There is a regrettably widespread assumption in philosophy that one can be a philosopher of mind without knowing what science has to tell us about the mind. We seem to have got over that sort of a prioristic isolationism in the rest of the philosophy of science (though I suspect some philosophy of language still goes on in splendid isolation from linguistics); it s time we got over it in the philosophy of mind. The idea that the philosophy of mind generally, and the theory of content in particular, is a branch of the philosophy of science, is not exactly universal. Perhaps it is contentious. So be it. I just can't imagine what else a respectable philosophy of mind could *be*.

Chapter 2

Contents and Targets; Attitudes and Applications

Beginning with Error

Theories of mental representation are frequently criticized for a failure to do justice to *mis*representation. Consider causal theories: According to causal theories, a typical case of *mis*representation is a case in which a horse causes a |cow|[1] rather than a |horse|. But if a horse causes a representation, then that representation cannot be a |cow|, since it isn't caused by a cow. Perhaps the representation in question is a |cow or horse|. But then it isn't a misrepresentation of a horse as a cow after all. There are cures proposed in the literature for this problem (dubbed the disjunction problem by Fodor), and for analogous problems that arise in connection with other theories of mental representation, but they are rather *precious*, hence not really convincing. There is more to good philosophy than being able to rebut the counterexamples.

Rather than begin with a theory of representation and then see whether it allows a coherent account of error, I propose to begin with a theory of error in the hope that this will place some useful constraints on further theorizing about representation. This may amount to little more than squeezing the balloon in a different place: one might pessimistically predict that it is bound to bulge *somewhere*. Still, a *new* bulge would be welcome. I'm tired of the disjunction problem.

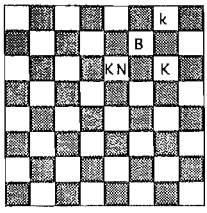
Targets and Contents

Suppose Σ is a chess machine. It uses a matrix, POS(row, col), to represent board positions. Σ is engaged in an endgame. The current position is P1 (figure 2.1), so

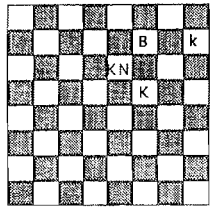
$$\mathrm{POS}(1,7) = k,$$

$$\mathrm{POS}(2,6) = B,$$

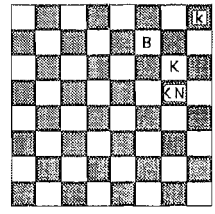
1. A |cow| is a mental representation whose content is (roughly) the same as the word "cow."



P1: the starting position



P2: the position after M



P3: the position R_{P3} actually represents

M=<k-kr1, K-KB4, k-kr2>

	1	2	3	4	5	6	7	8
1							k	
2						B		
3					KN		K	
4								
5								
6								
7								
8								

R_{P1}: the representation of the starting position

	1	2	3	4	5	6	7	8
1								
2						B		k
3					KN			
4						K		
5								
6								
7								
8								

R_{P2}: the representation of P2, the positon after M

	1	2	3	4	5	6	7	8
1								k
2						B		
3							K	
4							KN	
5								
6								
7								
8								

R_{P3}: a representation of P3, the representation Σ constructs to represent P2

Figure 2.1
Chessboards and their representations.

$$\text{POS}(3,5) = KN,$$

$$\text{POS}(3,7) = K,$$

and POS(row, col) is 0 for every other value of row and col. A subroutine, LOOK-AHEAD, anticipates *k-kr*1. To prevent stalemate, it investigates *K*-KB5, and anticipates *k-kr*2. Call this series of possible moves *M*. LOOK-AHEAD requires a representation of the position after *M*, viz. of P2. It requires R_{P2}. Suppose what it actually constructs, however, is R_{P3}, which represents P3, not P2. Since P3 is a stalemate, LOOK-AHEAD concludes that *M* leads to stalemate and rejects *K*-KB5 as a response to *k-kr*1.

Σ's tokening of R_{P3} in this situation is error. What makes it error is the fact that there is a mismatch between the state of affairs Σ needs to represent when it tokens R_{P3}, namely P2, and the state of affairs R_{P3} actually represents, namely P3. Call the former Σ's *target* in tokening R_{P3} on this occasion. Call whatever is specified by R_{P3}'s satisfaction condition—the condition that says that R_{P3} is satisfied by P3—its *content*. Then *tokening a representation is error when the target of tokening it on that occasion fails to satisfy its content.*

What makes P2 Σ's target in tokening R_{P3}? In the example, Σ constructs a particular data structure—the matrix R_{P3}—at a given point in the computation. What the computation "requires" at that point is R_{P2}, a data structure that represents the position P2, the position after *M*. That is, as it were, what Σ is trying to represent when it constructs R_{P3}, and that is what Σ proceeds to use that data structure to represent. Representing the position after M—P2 in this case—is, in short, the function of tokening R_{P3} on the occasion in question, even though, of course, R_{P3} does not represent P2 but P3.

The notion of a representational target is essentially a functional notion, then. When we say that the computation *requires* a data structure that represents the position after *M*, what we mean is that the function of tokening *R* (on that occasion) is to represent *T*. This formulation reflects the idea that we can explain *why Σ tokens R on a particular occasion* by appeal to the fact that *T* is *R*'s target. For example, Σ tokens R_{P3} in order to represent the board position after *M*, something it does in order to compute its next move. Note, however, that an appeal to *R*'s target does not explain why Σ tokens a representation with a particular content. In the example, Σ tokens R_{P3} in order to represent the position after *M*. R_{P3} does not represent the position after *M*, however, but P3 instead. To explain why, in this situation, Σ tokens a representation of P3, we need to investigate the actual computational history of Σ's production of R_{P3} on the occasion in question.

Targets, then, are determined by the representational function of tokening a representation on a particular occasion in a particular context, not by the content of the representation tokened. In our example, the target on the occasion in question is P2 regardless of what representation is tokened or what the content of that representation is. It is precisely the independence of targets from contents that makes error possible. If the content of a representation determined its target, or if targets determined contents, there could be no mismatch between target and content, hence no error. Error lives in the gap between target and content, a gap that exists only if targets and contents can vary independently. It is precisely the failure to allow for these two factors[2] that has made misrepresentation the Achilles heel of current theories of representation.

2. "Two factors": the target-content distinction should not be confused with the two factors in so-called two-factor theories (e.g., Block, 1986). The two factors of two-factor theories—typically conceptual role on the one hand and some causal or historical connection to the environment on the other—are designed to capture the distinction between narrow and wide content, not the distinction between correctness and error. A mismatch between narrow and wide content, for example, between water as colorless potable liquid and water as H_2O, or between brisket as breast of beef and brisket as breast of beast, is precisely NOT error, as Burge (1979), for example, is at pains to argue.

The obvious way to think of representational error, as Jerry Fodor once said to me in conversation, is this: Error occurs when a representation is *applied to* something it is not *true of*, for example, when one applies a representation of a horse to a cow. The distinction in this formulation between what a representation is *applied to* and what it is *true of* is precisely the distinction between a representation's target and its content. The crucial point is that what determines what a representation is *true of* must be independent of what determines what it is *applied to*, otherwise error is problematic. It follows from this that a theory of representational content—a theory that says what it is for R to be *true of* T—is only part of the story about representation. We require, in addition, a theory of target fixation, a theory that says what it is for R to be *applied to* T. Since the target of tokening a representation is, as it were, the thing the representation is intended to represent, I shall say that representations *mean* (or *represent)* their contents, but that a use of a representation *intends* its target. *Intentionality* is thus different from *meaning*; the former is part of the theory of targets, while the latter is part of the theory of representational content. Using this terminology, we can say that representational error occurs when there is a mismatch between what a representation means and what its use intends. The *intentional content* of r is therefore not the actual content of r at all, but, rather, the *intended* content of some use of r.

"Intend" is a technical term here. I do not, of course, suppose that cognitive systems generally intend their targets consciously, that is, that whenever t is the target of a use of r in Σ, Σ forms the intention to use r to represent t. But I do think the technical sense is a natural extension of this usual sense. In the case of conscious, deliberate use, intended content is quite literally the content one intends to represent. As always, one may not succeed in doing what one intends, hence one may fail to represent what one intends to represent. We will see shortly that the failure to distinguish intentional content (= target) from meaning (= representational content) derives from the fact that they are conflated in *attitude content.*

How do targets, that is, representational functions or intentional contents, get fixed? There will be an extended discussion of this question in a later chapter. For now, some rough indications and a few examples will, hopefully, suffice to give the idea. Think of cognitive systems as incorporating mechanisms whose function is to represent certain things. For example, the function of a simple visual system might be to represent the local spatial layout—the relative sizes, shapes, and distances of objects from one another and from the observer. When this mechanism constructs a representation, the target of the representation it constructs is the cur-

rent local spatial layout, whatever that happens to be. So the representational function of the mechanism, together with the current state of the world, determines what the current target is. I call mechanisms like this—mechanisms with specified representational functions—*intenders*, for the mind's contribution to intentional content, that is, to target fixation, is determined by the representational functions of such mechanisms. Perceptual systems are intenders in this sense, and they have more primitive intenders as components. The visual system, for example, might have intenders whose business is to represent corners or edges.

Simple (i.e., single-function) programming variables provide a transparent illustration of how targets get fixed in a computational system. A chess program might have a variable CURRENT_POSITION. Binding a value to that variable amounts to aiming a representation at the current position, because the program's logical design embodies the assumption that evaluations of that variable amount to looking up the current position. That is why we name the variable the way we do: We name variables after targets. The mechanism that constructs representations and binds them to CURRENT_POSITION is an intender: Anything whose function it is to bind representations to CURRENT_POSITION is an intender whose target is the current position. The "world," of course, determines (if anything does) what the current position actually is. So, as with the visual system lately imagined, the actual target of a particular use of a representation is determined by the representational function of the intender that tokens it together with the state of the world.

Evaluating CURRENT_POSITION amounts to looking up the current position regardless of what CURRENT_POSITION is actually bound to, that is, regardless of what the representation stored in CURRENT_POSITION actually represents. We can have error, and its characteristic consequences, precisely because CURRENT_POSITION can be bound to a value that does not represent the current position. The system will take the value of CURRENT_POSITION to be the current position, and, in that sense, the representation bound to CURRENT_POSITION could be said to represent the current position *to the system*, whether or not that representation is actually satisfied by the current position. There is a way of hearing the difference between "what R represents *to the system*," and "what R *actually* represents," as the distinction between target and content. But this is a dangerous way of thinking about the target-content distinction, because it tends to conflate it with the distinction between what a representation means *to* Σ and what it means *to us*. This latter distinction is not the target-content distinction, but the distinction between the content R has in Σ's representational scheme and the content it has in ours.

Intenders, then, not the representations in them, determine targets.[3] Indeed, given the way I have introduced intenders, saying that intenders determine targets is equivalent to saying that the target of tokening *R* is determined by the representational function of tokening *R*. A consequence of this is that representations are not, in themselves, erroneous. The same representation can be error in one intender (or in one intention, as we might say), but not in another. Consider Pavlov's dog Rover, conditioned to expect food, hence salivate, on hearing a bell. Hearing a bell reliably causes Rover to token a |food|, but this is not error because the |food| in question is tokened by an intender whose function is to represent the expected-stimulus. If hearing a bell were to cause an intender whose function is to represent the current auditory stimulus to token a |food|, *that* would be error.[4]

3. In earlier drafts of this manuscript, and in Cummins (1992), I used the term "BOX" for what I am now calling intenders. That use of the term "BOX" engendered a good deal of confusion because I wrongly introduced it as an adaptation of Schiffer's use of the term (1987). Schiffer's boxes are distinguished by cognitive function rather than by representational function, as intenders are. As Schiffer thinks of the matter, the belief box and the goal box are distinguished by such facts as this: The representations in the belief box are susceptible to retraction if undercut or rebutted by a defeater (Pollock, 1986), whereas, as every planner knows, a representation in the goal box should be retracted when the same representation appears in the belief box. What I am now calling intenders are distinguished by representational function. 'Belief' and 'desire' do not name intenders at all, not even very abstract ones, as I wrongly suggested in the works cited.

The ideas underlying Schiffer's boxes and my intenders are thus quite different. Schiffer wants to represent the difference between two different kinds of attitudes one can take to a proposition, that is, the difference between believing *p* and desiring *p*, the enabling assumption being the functionalist idea that this difference is a difference between two different ways of processing representations. I want to represent the difference between aiming *r* at *t* and aiming *r* at *t**. If that is what you are after, belief and desire boxes are non-starters, because the difference between belief and desire is not a difference in what you intend to represent, but a difference in what the system is going to do with the implicated representations whatever they are intended to represent.

4. The theory advocated here has the same form as that advocated by Dretske in "Misrepresentation" (1986). Dretske's formulation is this:

> *d*'s being *G* means$_f$ that *w* is $F =_{df}$ *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 its (*d*'s) being *G*.

This formulation creates a place for error to live, viz. in the gap between what something actually indicates and what is its function to indicate—its target, in effect. But note that what it defines is not the content of *d*'s being *G*, but the *target* of *d*'s being *G*. Or rather, it would define the target of *d*'s being *G* but for two factors. (1) The definition is in terms of indication rather than representation, whereas the function of most representations is not to indicate, as I shall argue in a later chapter. (2) The definition is evidently intended to apply to a representational type: the function of *d*'s being *G* is to indicate, that is, to covary with, *w*'s being *F*, and this makes sense only as applied to types, and it is tokens, not types, that have targets.

Falsehood and Error

It is tempting to think of misrepresentation as a case of representing falsely. We think of a system confronted with a cow and tokening a |horse|. That, surely, is a case of misrepresenting a cow as a horse. And that, in turn, is surely just representing a cow as a horse, that is, falsely representing something as something it isn't. It looks as if misrepresenting x is representing x, but falsely. The "mis" can just be dropped; it is not doing any work.

But this is a confusion of misrepresentation and falsehood. Once we see that targets are determined by intenders and not by the representations they token, we are in a position to see that error is distinct from falsehood. Suppose Σ tokens r when it needs to represent x as F, but that r represents x as G. This is error even if x *really is* G, so falsehood is not a necessary condition of error. Moreover, failing to represent x as F when x's being F is the target is error even if x *is not* F. Where is it written that a system needs only to represent the facts? Often—in hypothetical reasoning, for example—Σ will need to represent x as F even though x is not F. Falsehood in the goal box is normal; tokening truths in the goal box is a kind of pathology.

The point generalizes from truth to satisfaction. Most representations don't have propositions as contents, hence aren't the sorts of things that could be true or false anyway. R_{P3}, for example, represents a board position; it has a satisfaction condition, not a truth condition. What determines whether a representation like R_{P3} is error, however, is not whether there is some actual state of affairs that satisfies it, but whether it is satisfied by its target, which may not be an actual state of affairs. A target is determined by the function of tokening a representation on a particular occasion, in a particular context. Representational types, therefore, don't have targets, only tokens have them: they inherit them from the intenders that create them. Thus, a representation's content (its satisfaction condition), which attaches to the type, together with the state of the world, is not enough to determine whether or not tokening that representation is error. We need, in addition, to know the target of tokening the representation—what the representation is applied to—and that is determined by what intender has tokened it, that is, by the representational function of tokening the representation on the occasion and in the context in question.[5]

5. Of course, the content of a representational type, together with the state of the world, is not enough to determine truth value or satisfaction either, if there are, for example, indexicals in the representation. The present point is that fixing truth value or satisfaction won't determine whether there is error, since (1) some targets aren't actual, and (2) matching an actual object or state of affairs that is not the target is still error.

Since error isn't the same as falsehood, it follows that truth isn't the opposite of error. Indeed, truths are error when the targets are falsehoods, for example, when the target is a supposition in a reductio proof. Representations are often tokened to represent some particular false proposition, or to represent something that is not actual. Equally important is the fact that a representation of the proposition that p will be error when the target is the proposition that q, regardless of the truth values of p and q.

An embarrassing consequence of the fact that truth is not the opposite of error is the realization that we *have* no word for the opposite of error. I propose to call it *correctness*. A representation has been correctly applied when it hits its target. If r represents a proposition, then an application of r is correct if it hits the target proposition (assuming the target is a proposition). Since the target proposition can be a false proposition, the question of whether r is correctly applied is orthogonal to the question of whether it is true. If r represents an object or property, then an application of r is correct if it is satisfied by the target object or property. Since targets need not be actual, correctness is orthogonal to satisfaction by something in the real world.

Illustration of the Distinction between Falsehood and Error

There is, as I remarked at the beginning of this section, a temptation to think that we can do without the distinction between target and content because we are used to thinking of erroneous representation as false representation. For instance, we think of information stored in the mind in what psychologists call propositional form, that is, as sentences in the language of thought. Imagine a bit of stored knowledge:

1. Letters are more easily recognized in the context of words than alone.
2. In chess, one should develop the queen early.

We notice that (1) is true, whereas (2) is false, and this seems to be all there is to the issue of representational error: if we could explain where truth conditions come from, we could explain error: an erroneous representation is simply one that has an unsatisfied truth/satisfaction condition.

But this will not do. Consider (1). Letters are also more easily recognized in the context of pronounceable nonwords. Call the set of words and pronounceable nonwords the superwords. Then (3) is also true:

3. Letters are more easily recognized in the context of superwords than alone.

Imagine that a student is taught both (1) and (3) but stores only (1). Faced with a choice between (1) and (3) in a multiple-choice examination in which the question is, "Which of the following best describes the word-

superiority effect?," a student retrieves (1), yielding a wrong answer. Proposition 1, while true, is a misrepresentation of the target. It is error in spite of being true because its target is not simply The True, or a truth about letter recognition, but a particular truth about letter recognition. Thus, the truth conditions of (1) do not suffice to determine whether its "activation" or use on a particular occasion is error or not.

We can trace the error just imagined to the difference between "words" and "superwords". When (1) was stored, what was needed was an expression denoting the class of superwords, but what it "wrote" was an expression denoting the class of words. When (1) was constructed, the expression "words" was an error because the target of that expression, at the time it was constructed and stored, was the class of superwords, not the class of words. The case of terms, as opposed to sentences, makes it clear that a theory of truth and satisfaction conditions cannot provide a theory of error. Knowing the satisfaction conditions for "words" in (1) does not, and could not, tell us whether "words" in (1) is error. In addition to knowing which class of things actually satisfies "words," we need to know which class of things is the target of "words" in (1), and this, as we have seen, does not reduce to knowing whether (1) is true.

Representations, Applications, and Attitudes

Call an intender whose function is to represent t a t-intender. When a t-intender tokens a representation r, I call the result an *application* of r to t. For example, when the position-after-M intender tokens R_{P3}, we have an application to the effect that the position after M is P3. But the content of R_{P3} is not *that the position after M is P3*. To see this, we have only to note that when the current-position intender tokens the same representation, the result is an application with a different content, viz. *that the current position is P3*. R_{P3}, by itself, simply represents P3. This example shows that a system's applications can have contents distinct from the contents of its representations. In fact, the example shows more than this: the example shows that a system's applications can have contents distinct from the contents of *any* of its representations, for we may suppose that Σ is incapable of constructing a representation—an explicit data structure—whose content is *that the position after M is P3*. In general, then, one cannot infer that there is a representation with the content C from the fact that there is an application with the content C. When the position-after-M intender tokens R_{P3}, the result is a propositional application, but it is a propositional application that does not involve the explicit representation of any proposition.

We are now in a position to see how applications relate to attitudes. Propositional attitudes such as belief, desire, and intention are treated by

Schiffer (1972), Fodor (1990b), and many others as cognitive (or computational or functional) relations to representations, that is, as representations with a characteristic cognitive function. Σ harbors the belief that *p*, on this conception, if a representation that *p* is in Σ's belief box. But once we have the distinction between representations and their applications before us, it is clear that attitudes should be treated as relations to applications, not as relations to representations. In the chess example, Σ believes *that the position after M is P3*. But Σ harbors no representation, in its belief box or anywhere else, with the content *that the position after M is P3*. The only representation in the picture is R_{P3}, a representation whose content is a certain position, not a proposition at all. The propositional content of the attitude—belief in this case—is the content of an *application* of R_{P3} to a certain target. When the position-after-*M* intender tokens R_{P3}, we get an application with the content *that the position after M is P3*, the content of the attitude.

Attitudes are distinguished in type—for example, belief is distinguished from desire—by cognitive function. Since it is applications, not representations, that are correct or incorrect, it is applications, not representations, that have cognitive functions of the sort that distinguish the attitude types. It makes sense to retract an application of *r* to *t* when a defeater (e.g., an application of $-r$ to *t*) is discovered, but it makes no sense to retract a representation. What could it mean to retract a representation? Eliminate it from one's representational scheme? (One decides that one will never use four-letter words, or never use sentences of the form, "The bottom line is ..."). Evidently, the only relevant thing that could be meant by retracting a belief or desire is retracting an application of some representation, an application that is currently functioning as a belief or as a desire. Applications are the only kind of thing that can have the sort of functions that distinguish beliefs from desires.

An attitude is the result of giving a cognitive role to an application of a representation *r* to a target *t*. For example, a |bell| is applied to the target *current auditory stimulus* (|bell| ⊃ *cas*), and the resulting application is put in the belief box.[6] The whole business might be implemented by

6. In Schiffer's language, a belief that *p* is an application with the content that *p* in the belief box. Applications, however, are not the sorts of things that are naturally thought of as things you could put in a box the way representations are. One can put a (token) word or sentence or picture in a box, but not an application of a word or sentence or picture. Of course, talk of putting something in a box is simply meant as shorthand for giving it a certain cognitive function. But the shorthand is not apt: Applications can certainly be given cognitive functions, but they can't be put in boxes because they are events or states, not objects. Representations can be put in boxes, but they cannot be given the sort of cognitive functions that characterize belief.

1. Binding |bell| to the variable *cas*.
2. Making this binding available to defeasible inference but not to goal specification.

Notice that this application (binding) is now available for processing even though the system nowhere has a representation whose content is the content of the application |bell| ⊃ *cas*. The content that the current auditory stimulus is a bell is available even though it is not represented, because (a) the variable *cas* is implemented as a storage location, and (b) the system's logical design is such that the representations bound to *cas* are taken to be representations of the current auditory stimulus. Not only can you have a representation with the content that the current auditory stimulus is a bell without representing that proposition, you can *store* the attitude for future use. Of course, you don't want to have the belief that the current auditory stimulus is a bell for very long. But you won't. The system will write over the |bell| in *t* when the *cas* changes.

Can we get a standing belief like the belief that Socrates was a Greek philosopher without representing the proposition that Socrates was a Greek philosopher? In principle, this is simple. Implement a properties-of-Socrates intender as a variable *pos* whose bindings the system takes to be properties of Socrates, and bind it to |philosopher| and |Greek|. It is pretty unlikely, of course, that a natural cognitive system would implement such an intender directly. I don't know what directly implemented intenders humans have, but a properties-of-Socrates intender is surely not one of them. This is an empirical issue, of course. I don't think anyone has any well-justified ideas about what representational resources it takes for a human mind to have the belief that Socrates was a Greek philosopher precisely because no one really knows what the innate intenders are or how they are organized. What we do know is that you cannot read the required representational resources off the content of the belief.

Once we see that we cannot move directly from the premise that there is an application with the content *c* to the conclusion that there is a representation with the content *c*, we can see the mistake in the widespread theory that the content of an attitude is inherited from the content of its constituent representation. According to this theory, which Fodor (1975, 1980) calls RTM (Representational Theory of Mind), an attitude with the content *c* is a relation to a representation with the content *c*. This theory, which I will call RTAC (Representational Theory of Attitude Content),[7]

7. Fodor calls this the Representational Theory of Mind. In *Meaning and Mental Representation* (1989), I called it the Representational Theory of Intentionality (RTI), for I was thinking of attitude content as intentional content, and distinguishing it from the contents of representations. I still want to preserve the distinction between attitude content and representa-

licenses the inference from "Σ believes that *p*" to "Σ represents that *p*," an inference the discussion above shows to be incorrect. From "Σ believes that *p*" we can infer that Σ harbors an application with the content that *p*. But from the fact that there is an application *that p*, we cannot infer that there is a representation *that p*, as the chess example demonstrates. The RTAC, in effect, conflates application content with representational content. The RTAC thus leaves no room for error, for, by identifying application content with represented content, it leaves no room for targets to be specified independently of represented contents, hence no room for a mismatch between target and represented content. If I cannot distinguish the content of R_{P3} from the content of the application that results when an intender tokens it, then I cannot distinguish the content of R_{P3} from its target, and so cannot allow for error.

We can sum all this up in a few simple principles:

- Attitudes are applications with a characteristic cognitive function. The semantic content of an attitude is thus the semantic content of its constituent application.
- Applications are the result of applying a representation to a target. The semantic content of an application is that the representation hits the target:

$$\text{Application content} = f(\text{represented content, target})$$
$$= f(\text{represented content, intentional content})$$

As we have seen, representations in themselves are not correct or incorrect, though they may be true or false, satisfied or not satisfied; only the application of a representation to some target can be correct or incorrect. And that is why applications, but not representations, can have cognitive functions of the sort that distinguish the attitudes from one another, and hence why attitudes inherit their contents from their constituent applications, not from the representations involved. The semantic function of a representation is to represent its content; the semantic function of an application is to hit its target. An application of a representation *r* to a target *t* always has the content that *t* is represented by *r*. For example, applying R_{P3} to the position after *M* has the content *that the position after M is P3*. An application of *r* to *t* is correct when the content of the *application* is true, that is, when *r* represents *t*, and an application of *r*

---

tional content, of course: The content of an attitude always goes beyond the content of its constituent representation. But I now wish to reserve "intentionality" for the relation a use of a representation in an attitude bears to its target. Representations *mean* their contents, and a use of a representation *intends* its target.

to *t* is error when the content of the application is false. This, I think, is what gives rise to the idea that the theory of representational error and correctness simply reduces to the theory of truth. In a sense, this is right: representational error and correctness can be defined, as we just did, in terms of the truth or falsity of applications. But, since the content of an application is never the same as the content of the representation applied, a theory that gives us truth conditions for applications does not gives us representational contents.

Nesting Intenders

Consider the chess example once again, and imagine, as we did earlier, that POSITION_AFTER_*M* is a programming variable. *M*, we may suppose, is itself a variable that gets bound to representations of move sequences. Since POSITION_AFTER_*M* cannot be evaluated if *M* is not bound, the *M*-intender, that is, the thing whose business is to represent the move sequence of current interest, is nested in the position-after-*m* intender. Since intenders can be nested, and since you get an application when an intender tokens a representation, you can get nested applications. For example, we get an application with the content *that* $<k\text{-}kr1, K\text{-}KB5, k\text{-}kr2>$ *is the sequence of current interest* nested in a application with the content *that the position after the sequence of current interest is P3*: you bind *M* to a representation of the sequence $<k\text{-}kr1, K\text{-}KB5, k\text{-}kr2>$, then bind POSITION_AFTER_*M* to R_{P3}.

The phenomenon of intender nesting shows how it is possible for there to be an unlimited number of targets: intender nesting makes a system's targets both systematic and productive in Fodor's sense (1975, 1987). In the lately imagined case of the multiple-choice examination, we assumed that the target of,

> (1) |Letters are more easily recognized in the context of words than alone|,

was in fact,

> (3) Letters are more easily recognized in the context of superwords than alone.

Proposition (3) is the target of (1) on the assumption that (1) is tokened by a mechanism whose function is to represent the best alternative in the current choice set, or something like it. Such targets are obviously not a permanent feature of the functional architecture of the mind. They are created on the fly, perhaps by binding a lot of variables. Binding a lot of nested variables is a vivid way of seeing that computational systems can generate a wealth of rather definite targets as processing progresses.

When thinking about targets, it is important to keep separate two ideas about what fixes them. One idea is that (1) the target of tokening r is what Σ *needs to represent* to succeed. The other is that (2) the target of tokening r is what Σ expects to find when r is accessed. The first makes target fixation epistemic, the second makes it a matter of architecture or design. It is the second idea that I want to endorse. It is easy to see why the epistemic conception is problematic: there may be no content such that, were Σ to represent it, Σ would succeed, for Σ's success or failure will depend on many things other than the use of a particular representation on a given occasion. Moreover, it is easy to imagine situations in which representing t would bring success, but for reasons unrelated to, or inappropriately related to, representing t. Σ may succeed *because* it wrongly rejects a move whose consequences are incorrectly represented. Adopting the epistemic conception of targets expressed in (1) will make it impossible to articulate such possibilities.

The conception of target fixation I want, then, is the conception expressed in (2), that the target of tokening r is what Σ expects to find when r is accessed. Not literally "expects" of course; the idea is that Σ incorporates a design assumption to the effect that a representation generated by a certain intender will be a representation of t. This is, as remarked above, the conception that falls naturally out of thinking of intenders as mechanisms for binding programming variables. Just as programs are designed around assumptions about what will be accessed when a given variable is evaluated, cognitive systems are designed around assumptions about what will be represented by various intenders. Specifying a system's design or functional architecture involves specifying what intenders are possible, and hence, on the current conception, what targets are possible. The targets that are possible for a given system are thus fixed by its functional architecture.[8]

Can Representations Determine Targets?

The central point of the theory of error proposed here is that targets are determined independently of representational contents. The question arises, therefore, as to whether a representation could somehow specify its own target.

In the abstract it seems clear that this cannot happen. T is the target of a given tokening of R just in case the function of that tokening of R is to

8. We will have occasion to qualify this in a later chapter, for, as mentioned earlier, there is often an indexical element to target fixation. That there is a current-position intender in Σ is fixed by its architecture. But what the current position is, and hence what a representation generated by the current-position intender must match to be correct, is determined by the world.

represent *t*. But the function of a given tokening of *R* is surely independent of the content of *R*. I don't mean to claim, of course, that the function of *tokening something with the content c* is independent of the content of *R*. Contents, no doubt, have functions, as Millikan (1984) has emphasized. But we must distinguish the function of a representational type—something that plausibly turns on its content—and the function of tokening it on a given occasion, which does not. The function of R_{P3} is, I suppose, to represent P3. But the function of its tokening by the current-position intender is to represent the current position, *whatever that is*. Since the current position need not be P3, it follows that the function of tokening R_{P3} cannot, in general, be the same as the function of R_{P3}.

But couldn't a representation say, in effect, "The Eiffel Tower is my target"? Certainly. But it won't follow that the target of this representation *is* the Eiffel Tower. Indeed, the fact that this representation says that its target is the Eiffel Tower will have no effect whatever on what its target actually is. To confirm this, we have only to remember that the representation in question could be generated by any number of different intenders. For example, an intender in the language-understanding system whose function is to represent the meaning of the current sentence might generate some such representation. Your language-understanding system did just that a moment ago. When it did, the target was not the Eiffel Tower but the meaning of the then current sentence, and the determination of that target was quite independent of the content of the representation you tokened.

Target and Referent

The fleeting temptation to suppose that the content of a representation could influence the target of some particular tokening of it derives from a tendency to conflate targets and referents (and, correlatively, content with sense): if you think targets are just referents, then, since a representation can specify its own referent, it seems it could specify its own target. Consider an |I am rich|: Isn't the target here the "speaker"?

To see what is wrong with this, imagine a context in which what I need is a representation of the proposition that Scrooge is rich, that is, a context in which my target is *that Scrooge is rich*. If what I produce on that occasion is an |I am rich|, we have representational error whether or not I am rich (provided I am not Scrooge, of course). The target of "I" need not be the speaker—that is its content. The target of "I" could be anyone or anything. That is what makes it possible for a tokening of "I" to be an

error.[9] It is pretty obvious, on reflection, that targets are distinct from referents, for otherwise reference could not go wrong. Imagine a representation *R* whose referent is my right hand, and an intender *L* whose target is my left hand. When *L* generates *R*—the sort of thing one's visual system is liable to do when one is looking in a mirror—we have error. It is error precisely because there is a mismatch between referent and target. Satisfaction conditions—conditions that say, among other things, what refers to what—are contents, not targets.

It is easy to get confused about this if we are thinking about the content of an application rather than the content of a representation. The content of an application is, as we have seen, contributed in part by the intender. The possibility thus arises that the referent of an application is contributed by the intender. Since intenders determine targets, applications can refer to the targets of tokening their constituent representations. When R_{P3} (i.e., |P3|) is tokened by the current-position intender, the result is an application with the content *that the current position is P3*, that is, an application that is about the current position. Here, what the *application* is about—what it refers to, in one straightforward sense—is its target, viz. the current position, P1. But what the *representation* is about—*its* referent—is P3. Overlooking the distinction between the content of an application, and the content of its constituent representation, therefore, can lead to a conflation of the target of tokening *R* and *R*'s referent.

Three Theories of Mental Content

We need, then, three distinct theories to deal with mental content:

1. A theory of representational content.
2. A theory of target fixation.
3. A theory of application content.

The *theory of representational content* must explain what it is for something to be a representation, and what it is for a given representation to have a particular content. A crucial constraint is that representational content should be distinct from application content and independent of target fixation.

The *theory of target fixation* must explain "the function of tokening a representation *r* is to represent *t*." A crucial constraint is that target fixation must be independent of representational content.

9. The content of an |I am rich| is, in any case, a proposition, not a person. When an |I am rich| is correctly applied, its target is the same as its content, viz. the proposition that I am rich. If we suppose that the target of an |I am rich| is the speaker, we are bound to get error, for then the content, being a proposition, has no chance of matching the target, a person.

Since the content of an application is a function of two factors, its target and its representation, the *theory of application content* must explain how target and representational content add up to the content of an application.

The problem of mental content is, then, three problems rather than one. Only by keeping these distinct is there any hope of emerging with a conception of mental representation that makes room for a robust conception of error.

Some Diagnostics

Why has the crucial concept of a target been neglected? Here, I think, is how it usually happens. One begins with an example such as, "Σ believes that the position after *M* is P3." That belief is error, one supposes, just in case it is false that the position after *M* is P3. Error in belief seems to be just the falsehood of a representation whose content is specified in the that-clause. One's theory of truth, then, will do as a theory of error. All that remains is to explain what it is for the belief to have the content *that the position after M is P3*.

The trouble starts when we consider an example divorced from any particular processing context. When Σ comes to believe that the position after *M* is P3, Σ is not simply in search of some truth or other, it is after the position after *M*. To get things right, Σ needs a very particular belief, a belief with the content that the position after *M* is P2. Realizing that another truth won't do is enough to show that a theory of truth won't do as a theory of error. One needs, at a minimum, to ask which truth the system is after on the occasion in question, and that is to ask after the target.

The minimum, however, isn't nearly enough. The story above moves immediately from the premise that Σ has a belief with the content *that the position after M is P3* to the conclusion that Σ harbors a representation with that content. Attention to the details of realization and processing reveals that this is a faulty inference. What Σ needs, on the occasion in question, is a representation of P2 targeted at the position after *M*. The only representation that enters the picture is R_{P3}, a representation of a position. We thus arrive at the distinction between the content of an application and the content of its constituent representation. Since beliefs with the same content can be differently constituted by different target plus representation combinations, and since targets are fixed by intenders, it follows that two beliefs with the same content can have different targets. When the contents-of-the-basket intender generates a representation of the cat, you get a belief with the content that the cat is in the basket, but the target is the contents of the basket. When the location-of-the-

cat intender generates a representation of the basket, you get a belief with the content that the cat is in the basket, but the target is the location of the cat. The conditions for error are, therefore, very different. A representation of the cat, however excellent, is error in the latter case, while a representation of the basket is error in the former.

At least as serious as ignoring the details of processing and realization is the obsession with belief to the exclusion of the other attitudes. Truth is evidently not typically a necessary condition of correctness in intention or desire, and this obvious fact should have been enough to generate a distinction between error and falsehood. Unfortunately, it is all too easy to drop the ball by supposing that a correct intention or desire is simply a rational intention or desire, and ship the problem off to epistemologists without noticing that the subject has been changed from representational error to inferential error.[10] Since inference is something philosophy is used to understanding in terms of truth, it might seem that error in intention and desire really introduces nothing new.

The bottom line is that you cannot hope to do justice to mental content by thinking about the truth conditions of belief attributions. The philosophy of psychology isn't the semantics of folk-psychological locutions. To certain philosophers of mind this may seem like an unemployment opportunity. But, after all, it really never was in the cards to get a theory of mental content out of a semantics of English belief sentences (Cummins, 1991a).

10. This is a change of subject that conceptual role semanticists are bound not to notice, as we will see in a later chapter. Since conceptual role semantics identifies *R*'s content with *R*'s epistemic liaisons, something with a "bad" content is just something involved in some bad inferences.