

Putting One's Foot in One's Head
—Part II: How

Donald Perlis
Department of Computer Science
and
Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742
(301) 405-2685
perlis@cs.umd.edu

April 22, 1992

Chapter 1

Putting One's Foot in One's Head—Part II: How

1.1 Abstract

This is the second of a two-part work on intentional states. My thesis is that an *internal* notion of reference is far more salient for philosophy of mind and language, as well as for artificial intelligence, than is the usual external one, and that the former provides a key to the latter. This involves in a central way the embedding of the thinking agent in a causal physical environment, and thus has much in common with functionalist views. However, it also involves a self-referential feature that can be brought to bear on the subjective (narrow or individualist) character of thought, as well as on the problem of error, and which provides the aforementioned key between internal and external reference.

In *Part I: Why* of the broader work, I argued the need for (and suggested an account of) internal reference, as part of any successful theory of external reference. Here I take up issues left aside in *Part I*. In particular, I focus on *how* a meaning agent can accomplish tasks of external reference (largely via internal reference). My proposal is a combination of several others well-discussed in the literature—covariant, historical, imagist, and description-theoretic—plus some new features (involving self-reference and some modest insights from neurophysiology).

Contents

1	Putting One’s Foot in One’s Head—Part II: How	1
1.1	Abstract	1
1.2	Introduction	2
1.3	Some related ideas	4
1.3.1	Dealing with Putnam’s internal realism	4
1.3.2	Fodor’s long way ’round, and thought-ascription	6
1.3.3	Qua problem	7
1.3.4	McCarthy Machines	7
1.4	Body-reference	8
1.4.1	The inner metric	9
1.4.2	A difficulty	11
1.4.3	Some neuroscience	11
1.5	Beyond the body	12
1.6	Error and recovery	15
1.7	Functional role semantics vs. iref	17
1.8	Reflection and geometric semantics: a summary	18
1.9	Virtual reality machines	21

1.2 Introduction

In the “prequel” (*Part I* of the broader work), much was said about *why* it is important to have an inner world model, including a model of the external referential relation itself, in any entity that purports to have intentional states. However, little was said there as to *how* this might be accomplished, i.e., how it contributes to an account of *external* reference. In this current paper (*Part II* of the longer work) I will address this matter, and propose a specific architecture toward that end. This architecture should be viewed as an addition

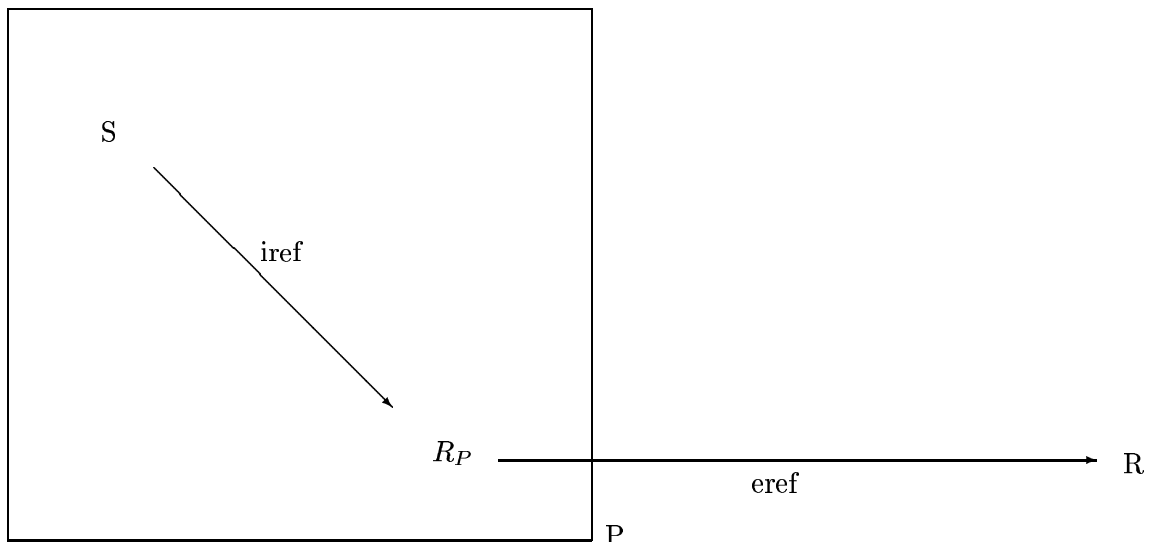


Figure 1.1: Two-Tiered Model with Iref and Eref.

onto the familiar functionalist architecture, and not a departure from its basic conceptual framework.¹

I will attempt to provide an account of external reference along the following lines: An external object R is related to the meaner (reasoning agent) P in much the same way that the notional object R_P is related to P 's self-notion P_P . Thus we introduce a self-notion in part for this purpose. The basic (easiest) case is that in which R is part of P 's body (e.g., R is P 's foot). Then the “in much the same way that” allusion two sentences above can be given physiological force via the actual pathways between foot and brain (e.g., leading to the tectum). Once we have a semblance of body-meaning along these lines, we can try to extend it further outward by a kind of “body geometry”. (In this we are not far from the spirit of (Sloman 1986), (Steels 1986), and (Johnson 1987); see also (Perlis&Hall 1986), and (Perlis 1986, 1987, and forthcoming)).

First, I remind the reader of some of the main points from *Part I*, concerning iref (internal reference) and how it is intended to fit into a larger framework of eref (external reference). In Figure 1.1 is shown a sketch of the theory of internal reference given in *Part I*. There the symbol “ S ” is some internal token, which (we wish to explicate) somehow has content R external to the reasoning agent P . This occurs in two parts: iref maps S to another internal entity R_P , which is the reasoner's version of R (i.e., in the reasoner's world view or scene), and R is the actual external object that R_P corresponds to (via a relation called eref). I will not here go into why this internal architecture is proposed; this is dealt with in detail in *Part I*.

The point is that, in order for the reasoner P to refer (as opposed to merely intone, say) when she says “ S ”, she must have something in mind that she means (in *Part I* this is called the Generalized Kripke Condition). The latter thing-in-mind is R_P . Yet for this to have any external content, there must actually be some external thing, R , that S and P and R_P

¹ Although functionalists must decide for themselves whether they regard my additions as a basic departure or not!

somehow are suitably related to. The remainder of this paper is devoted to an account of this latter relationship (eref). My account will treat eref as dependent upon iref in a crucial way. The upshot will be an account that can be given the following formal dress:

$$\text{content}_P(S) = \text{eref}_P(\text{iref}_P(S)) = \text{eref}_P(R_P) = R$$

A question that naturally arises is what good is the extra layer of tokens; why not just as well go right from S to R ? There are several reasons for this, discussed at length in *Part I*. But for our present purposes, it is especially important to recall one of them: error recognition and correction. If a reasoner is to be able, at times, to come to believe that a prior belief (B) of his may have been in error, then he needs a way to represent that belief as an object (a former belief) and what it was about (A), and also the new believed information to the effect that there is a discrepancy between A and B: B is (contrary to prior belief) not true of A. In terms of our above notation, for instance, if S is the word “dog” then the belief that R_P is a dog may later come under suspicion and be replaced by a belief that R_P is really a fox but appeared to be a dog at first. This is treated in a later section below.

1.3 Some related ideas

1.3.1 Dealing with Putnam’s internal realism

Putnam [1975] argues that objective accounts of meaning must fail, and he proposes instead an “internal realism” that seems to correspond to the “subjective inner view” that each individual has, as we might say, of “the world”, although Putnam is denying that there is any one correct match between these inner views and the outer world. As a consequence, says Putnam, our words can never be (in any robust sense) genuinely about external things. We can refer only to internal things, things in our views.

If Putnam is right, we can’t be a brain at all, not even in a body! For our internal meanings do not pertain to things outside, such as our bodies out there. Our notion of body is “in here” in the sensorium, the mental space our minds inhabit. So to us, according to Putnam, “body” cannot mean something out there beyond our thoughts; all our thoughts mean things in our world of thoughts. This is much more than the mere thesis that there is a *puzzle* about how internal things like thoughts get tied to external things like tables and chairs. The latter is a recognized hard problem, sometimes called Brentano’s problem, and is usually taken as something to try to solve. But Putnam is presenting an argument that purports to show there *can be* no solution to Brentano’s problem: it is folly to think our thoughts get mapped outside in any coherent, objective way, he says. This is a very radical thesis. It seems to require giving up the standard scientific view of an objective external reality about which we think.

What does our iref/eref distinction have to say about it? We will need to elaborate our account of eref before a clear statement can be made, but in rough form it is as follows. Putnam’s argument involves some considerations that apply to iref, and others to eref. His argument uses the following criterion that is supposed to apply to any possible theory of meaning: the meaning of a sentence depends crucially on the meanings of its parts (words). That is, we cannot in general alter the words (or their meanings) without altering the

meaning of the whole sentence (in at least some way). This criterion is sometimes called “compositionality”. Putnam argues that in fact nearly *arbitrary* changes in any purported external word meaning can be allowed and yet preserve sentence meaning, thus violating compositionality.

Now, iref does obey compositionality, and it does indeed provide a way to link symbol to symbol. If we make a sentence using symbols, the (iref) meaning of the whole will reflect the (iref) meanings of the individual symbols. However, the (iref) meaning here is determined by the user (this is what iref amounts to) and so any change in word (iref) meaning will in general change the overall internal user meaning of sentences using those words. This is so since iref maintains the reference relation in an internal structure, not in some unreachable external world. If the meaning of “foot” is changed, it is changed because the user has so changed it, i.e., she has decided to now take “foot” to refer to, say, her hand (as construed by her in her inner view of her limbs).

Where Putnam’s claims have force is in the case of eref. If we readjust the agent’s world so that in fact $eref_P(\text{foot}_P)$ is no longer the agent’s foot, and similarly for $hand_P$, say, it is possible *a la* Putnam to get a sentence such as “my hand is on my foot” to have the same eref-ential meaning as before the readjustment. But the readjustment is a matter of engineering the way the agent fits into the world (e.g., rewiring the agent’s nervous tissue between brain and limbs), and does not in itself affect what the words mean to the agent (i.e., does not affect iref). This may be in part Putnam’s point: (the compositional part of) reference is internal. But he ignores the possibility of a robust external component that allows for an objective external referent after all, but which is not (and has no need to be) compositional. Eref accounts for what links between (internal) mind and (external) world make the former’s behavior fit well into (survival in) the latter.

What an external account is, then, is evidential. We do not pin down an external (eref) account for certain. We make educated guesses, and rely on evidence to provide us with more or less confidence in our theories about externals. I will provide below an external account that is in rough accord with commonsense (I claim), but certainly is not to be regarded as an *ultimate* determination of external reference. Indeed, there may well be no such ultimate determination, for our concepts may in large measure be usefully related to the external world only at a special human-sized scale. But this is no mark against such an account of meaning. We can indeed mean tables by “table” even though a sufficiently detailed account of matter may convince us that instead of “real” tables there are gyrating conglomerations of atoms held together loosely by potential force fields, that merely look table-ish to us. Certainly the modern notion of tables has undergone considerable change from older views of tables as solid homogeneous masses. Yet people back then surely did refer, as much as we do now, by the word “table”, in the sense that their behavior calls for an explanation pretty much the same as ours. The discovery of atoms did not, surely, suddenly and totally alter the basic nature of language and mind.

We form theories as to an external world. These theories are internal. We revise them as we get more data. This is true whether in everyday life or in the technical realm of science. The revision process is a complex one, in which we “reflect” our own processes into a more subtle representation of our thinking. The extent to which an external reality may or may not actually correspond to our thoughts is not one we have means to ascertain except tentatively by a kind of educated guesswork. However, we do have by our own lights a

variety of tools to decide if our theories are adequate in terms of what we know and expect of them and our data. Moreover, these tools include measurements, and measurements implicitly involve interaction with the environment. It is this interaction that is the basis for the theory of external reference (eref) that I will develop below.

Relative to a viewpoint, we can determine eref and hence content. This is Putnam's point, in a way. But we alter our viewpoint as we get new data—the viewpoint is in part honed by us to be useful, succinct, to fit the data into packages we can deal with. We can even often get away with the pretense that there is a single stable scientific viewpoint within which we carry on our theoretical discussion of mind and language. For when that viewpoint goes awry, we have worse things to think about than intentionality.

1.3.2 Fodor's long way 'round, and thought-ascription

Fodor has been working long and hard on Brentano's problem, insisting on a rigorous mentalism: he takes beliefs and other cognitive states seriously, thinks they are real indeed, and also that they nevertheless are quite naturalistic events, ultimately explainable in non-mentalist (and objective) terms. However, his efforts also tend to have large doses of distal-ity in them. Instead of looking squarely at behavior, he looks elsewhere to get a side view through counterfactual glasses. And he bases this in part on the distal theories of Putnam, Kripke, and Devitt. By "distal" I mean an account that focusses mainly on external reference, with relative neglect of internal reference. By contrast, a proximal approach looks more closely at internal workings of the mind, and in particular at the internal referents of symbols.

There is a clear appeal to distal approaches. We know little about the mind *per se*; and many different mental operations would seem to be able to be "about" the same external object—so different in fact that it seems clear no one account of the mental could possibly capture cleanly all kinds of thoughts about ducks, say. To understand this, we need to further break down duck thoughts into two types.

One type of duck thought is ascribed duck-thought: we say of someone else that she is thinking about a duck. This even though she may not know the word "duck", nor even have a concept at all similar to our duck concept. She may, perhaps be thinking of a particular bird flying overhead, that happens to be a duck (unknown to her). we still can fairly say that she is thinking about a duck, but nothing in her brain (or mind) has anything to do with a duck-concept (we shall suppose). So far so good. In this "ascriptive" sense, it will be impossible to give an internal (individualist) account of which thoughts are duck thoughts. However, this I think is not what we really want to do anyway. Ascribed duck-thoughts are things for our benefit, and are not very revealing about the person so ascribed as having duck-thoughts. This is the same point I make in *Part I*: if we want to understand P's behavior, it is P's view we need to analyze, not ours.

The other kind of duck-thought is one that is a duck-thought to P herself. This is determined by her (iref) use of the word "duck". And here there seems every reason to believe that all thoughts that are duck-thoughts to P, have something in common in her brain. After all, they are all identified by her as "duck-thoughts".

1.3.3 Qua problem

The ‘*qua*’ problem is as follows: does “foot” mean my whole foot, or just part of it? (This also is Quine’s famous problem of indeterminism.) But as far as iref is concerned, there seems to be an easy answer. “Foot” means whatever is actually tagged $foot_P$ by the tag “foot” in my mental space (i.e, tagged by the iref relation). This sheds some light on iref: the tagging should be viewed not merely as a matter of tying a string to R_P with a card at the other end having the tag S written on it.² Rather, the entire taggee R_P should “light up” when the S tag (or button?) is pulled (pushed?). That which is lighted is then the whole internal referent of the tag.³ Now we can then extend this to external reference (eref) by the same means as before, letting R_P determine via body-geometry that part of real external space-time that is the externally corresponding entity.

Now, this is an easy case, for we have been dealing only with singular terms (names). What about “feet” as a kind term? Does it apply only to feet, or to any bodily appendages? Again, we can tell by seeing how it is used internally with respect to parts of the inner homunculus. And again, the meaner (P) has control over this. She can decide to use “feet” to refer to appendages if she wants. Then not only $foot_P$ but also $hand_P$ will light up in her mental space whenever the “feet” tag is invoked.

1.3.4 McCarthy Machines

McCarthy has offered an example to illustrate the precarious nature of intentionality.⁴ I will present the example and then turn it around to argue against the Dennett-McCarthy intentional stance.

McCarthy’s example (more or less): Suppose that a machine is constructed so that whenever one or more persons are nearby it does a quick scan of their DNA pattern (with futuristic technology) and if that pattern matches a fixed one that had been randomly selected and kept in its database years ago then it loudly intones “Get him!” and releases a robot that grabs any body present with that DNA pattern. Suppose further that the stored DNA pattern is exactly mine and no one else’s. Is it the case then that the machine’s intoning “Get him!” means me? Is the term “mean” here a significant theoretical concept or a bit of fluff that we can dispense with? I think in this example it is fluff. All that is significant, all that there is to the matter, is that if I do not remove myself quickly from the neighborhood, I will be captured. The machine is a DNA-pattern recognizer with a subsidiary grabber: that tells the whole story. Whether we call that meaning or not is immaterial. It tells us nothing new to say that the machine’s use of “him” is symbolic of me. The intonation results from my DNA triggering a rather dull sequence of actions,

²For then the *qua* problem would resurrect itself: what is the extent or boundary of the thing at R_P ’s end of the string? all of R_P or part of it, or more than it?

³We should avoid too narrow a reading of “lighted up”; this is not a visual image in any naive sense. For instance, we may envision or think of R_P as a solid object entirely lit up on both its surface and its interior. What matters is what we take to be referred to; we apparently have a great skill at imagining various things, and that is what lets us think so subtly.

⁴personal communication. In particular, McCarthy believes his example shows, with Dennett, that whether an agent’s behavior exhibits intentionality is a stance we, the observers, take, and is not a matter of fact about the agent herself.

that is all. McCarthy and Dennett, I think, would agree with this. We can say that the machine means me, if we wish; it is convenient to talk that way, but reveals nothing about the machine in itself. On the other hand, if my identical twin comes by, the machine will treat him the same as me, so it seems misleading to say the machine really means me, specifically, by “him”.

However, there are cases in which it is not immaterial at all, but central to the events, in fact needed in order to understand what is happening. These are cases of the genuine use of symbols as symbols, i.e., for the purpose of standing in for something else. Now, how do purposes get in here? And what good are stand-ins? MacBeth’s dagger is a good example. The dagger-percept is not a dagger, and he knows this. What he wonders is whether there is a real dagger that is giving him the dagger-percept. That is, he knows the percept is not the real thing, whereas McCarthy’s machine has no such information or ability to process it (unless it is constructed so as to have these, in which case it then is using symbols as symbols).

Now, this sounds similar to the original McCarthy machine. It too does a comparison of input DNA-pattern to stored DNA-pattern. The difference is in whether any relating is done to a potential external entity, i.e., whether the machine has a world model in which its own processing is represented as part of that model and potentially external things as other parts that can have stand-ins inside its self-model. Only then is it representing a stand-in relation, i.e., a meaning relation, a difference between its percept (the input DNA-pattern) and what it takes to be the external cause (me, say). Yet to fulfill McCarthy’ original scenario, none of this sophistication of double representation—of separating the thought or symbol from the think thought about or symbolized—is needed. And for the same reason, that machine cannot question whether it got the wrong person: it has no mechanism to distinguish people (or DNA) from its data-patterns that we would regard as being about people (or DNA). If the pattern matches, then that is that; there is no notion, for it, of there being an error in the scanner that read the DNA-pattern, i.e., that the real DNA being scanned has a different pattern from that which in input. It has nothing with respect to which there can be an error: no inner model providing what for it would be a correct outcome that can fail to occur.

On the other hand, the more sophisticated version of the machine might indeed be able to detect what it would regard as errors; for instance if the robot cannot grab anyone despite the pattern-match, it has the internal representational material to formulate the notion, for instance that there is no external cause of the pattern, that it is an illusion; or that the external cause is nevertheless not the person with the indicated DNA-pattern, due to processing error.

1.4 Body-reference

In this section I present the bare bones of my theory of *eref*, in a restricted case, namely, when the “external” referent of a token is part of the reasoner’s own body. The reason for this is as follows: it is tempting to say that the relationship between R_P and an external R is essentially that between S and R_P : S is to R_P as R_P is to R. This is especially suggestive in light of our insistence in *Part I* that the reasoner, P, must have an inner model of herself and her surroundings, if she is to have beliefs about these. For a picture, see Figure 1.2

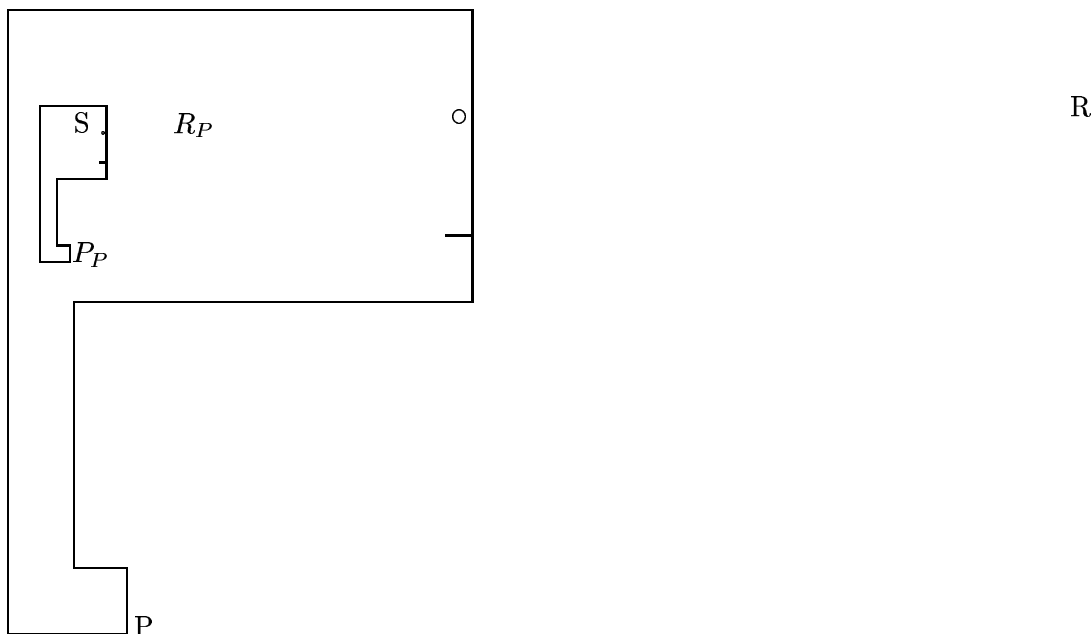


Figure 1.2: One's inner model.

Here the initial idea is to say that P_P is to R_P as P is to R . The trouble is that there is no obvious metric by which the “ P_P to R_P ” relation is given. Our task is then to find such a metric. This appears to be easier when R is itself a part of P 's body. Then later we can see how this might be extended beyond the body to distant objects. But even when R is a body-part, how can we explicate an inner relation that is like an outer (bodily) one? Pictures surely will not do. While a definitive answer will be a true feat of neuroengineering, I think that at least a plausible outline can be developed. And, in a way, pictures are the key. For whatever our minds do, they can conjure up in the mind's eye inner representations that we use to get around with.

1.4.1 The inner metric

We will take one simple example. Suppose the word “foot” plays the role of our symbol S above. Then R will be P 's actual foot, and we seek an account of the relation between these two, via P 's inner foot-notion, what we might write as $foot_P$. Again a picture is instructive; see figure 1.3. $foot_P$ is a sort of “footunculus”, to be discussed more below. We thus focus on a situation that is hard-wired into our bodies, namely, self-perception. Consider what it is to refer to one's foot. One has the name “foot” (or whatever or however one chooses to call or token it) and also the bracketed foot, that is, the thing we take (in our mental space) to be the foot. Now, the former, the symbol (say “foot”) is under our control. We can instead employ “pied” or “shmoot” or anything we like. But our bodies/brains provide us with a hard-wired footunculus, that is created as a result of wiring between the brain

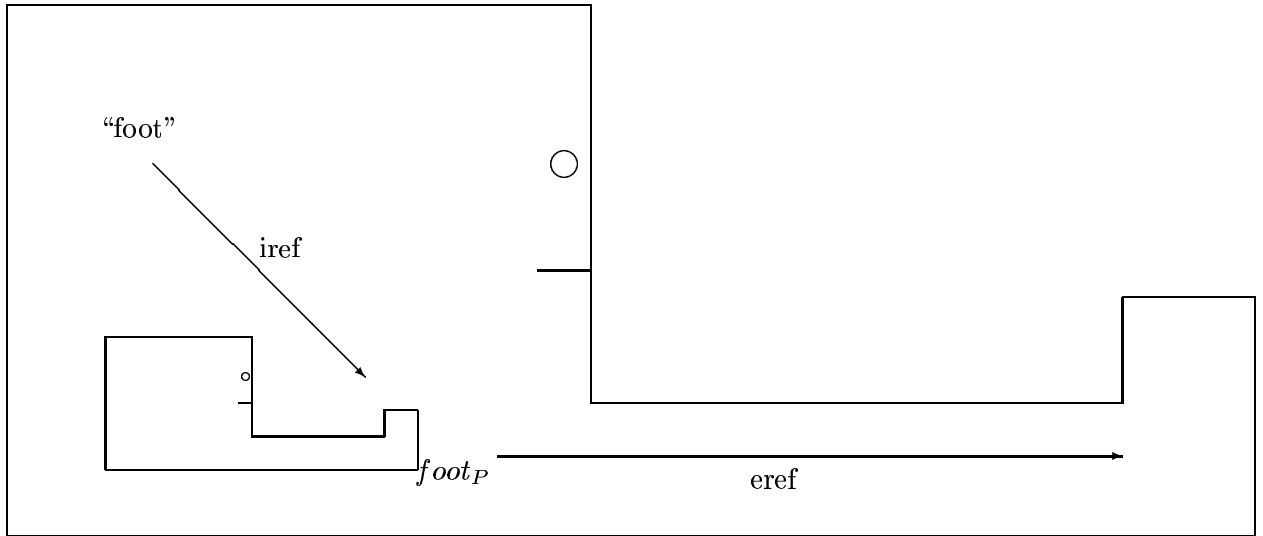


Figure 1.3: The foot in one's head.

and (you guessed it) the actual foot extremity itself.⁵

Note that $foot_P$ is not the word “foot”, nor is it the equivalent mentalese of “foot”. $foot_P$ is the visio-proprioceptive “reliable sign” of P’s foot in P’s head. This latter tie is not $iref$, rather it is $eref$. Note also that not merely feet but much of one’s body is so mapped via the nervous system to an inner homunculus (actually several homunculi, which we ignore here). These inner bracketed body parts can then be named by us (attended to as “this part” and “that part”) and thus thought about. Whenever we “think about our foot” we are invoking or activating a relation between our footunculus and some name or other ostensive tag that—among other things—declares our attention to it. It is the ostensive tag that is the symbol; we can use whatever we like as tag, as long as we make sure that we in fact so use it, i.e., we tie it appropriately to whatever we want it to tag. This amounts to a procedure that we create and invoke: it takes us from the tag to the taggee, and our job is then to remember the procedure (i.e., remember which tag goes with which taggee).

This footuncular map may obtain even when we are not looking at our foot, but merely thinking about it, for the map is not solely tied to the visual system.⁶

The internal metric that provides the congruence between the S-to- R_P relation and the R_P -to-R relation, is simply that given by the neural wiring: whatever internal structures allow us to make assessments of, say, the length of our foot, determine the metric. The “length” of our footunculus is (mapped to) the length of our foot. If there is a discrepancy, this will cause us problems, which we either learn to correct, or which leaves us in some confusion and without a sharp semantics for, say, length. It is easiest to suppose an actual inner body map, much like our actual bodies, although this is physiologically mis-

⁵In fact, there are several such footunculi in various parts of the brain. For instance, the tectum (part of the mid-brain) involves so-called tectal maps, which facilitate computing spatio-temporal trajectories based on inputs from visual, auditory, and proprioceptive systems. We need the footunculus and its motion in our mental space to reliably covary with real foot motion.

⁶For the present I ignore difficult cases such as that of an amputee who has no foot, or of someone who has a foot but whose neural wiring does not provide an inner footunculus. I return to this later.

leading. Some other story, but presumably not wholly unrelated, remains to be uncovered by physiologists.

Time here is not a symbolic/formal notion of time, it is real everyday time. We must have a genuine clock that we can rely on to show us this is *now*, that was *then*. We also must have a genuine space or field of tokens we can employ for tags, and be able to tell them apart, and keep track of how we have decided to hook them to inner referents. This too must be real, actual physical structure that we can recall and shape to our needs. Thus the mental engines of thought are fundamentally embedded in a physical context and rely on accurate access to the context to inform of present versus past and of this tag versus that.

There is then a primitive space-time semantics built into the inner ostending machinery. It distinguishes times, places, tags, and taggees. And this, I hope, is enough to regain the world as we know it.⁷ It must be—modulo alterations of letter but not of spirit—if my position is to stand.

1.4.2 A difficulty

Now we address the situation of the amputee. On a naive reading, my account so far would seem to say that if I lose my foot, then I lose meaning for the word expression “my foot”, for I am no longer appropriately wired to it. I may still have my footunculus *foot_P*, but this is mere iref. Yet our intuition is that I can still refer to my missing foot. Partly this is a problem of extra-bodily eref, that we will get to in the next section. But since I will build those treatments based on bodily eref, how can my theory hold up in the case of an amputee?

Here we are to suppose that one’s foot-reference is not the only semantic tie one has to one’s body. Suppose we also have ties to hands, fingers, arms, eyes, etc. Then if we lose a foot, we can still interpret our bodies via these many other intact regimes, and from there try to get on to full extra-bodily eref. On the other hand, a creature with only a footuncular inner self would indeed, on my theory, lose its semantic attachments when it loses its foot. By the same token, if we lose all our homunculi-to-body wirings (super-paraplegic) then we do lose our semantics as well. Without eyes, ears, limbs, skin, etc, I cannot relate my inner body notion to a body (for there is no body to speak of, and I am just a brain in a vat, indeed an unfriendly vat in which I do not get the right sort of inputs when I perform certain output commands).

1.4.3 Some neuroscience

It is instructive to ask whether the internal layer of representation I am espousing has any basis in biology. Certainly we do not, in any literal sense, have little people in our heads. However, it is no joke that a sort of homuncular view of the mind/brain is central to much of modern neuroscience. There are in fact a number of distinct homuncular maps between the body overall and portions of the brain, whereby certain brain structures correspond (in the

⁷Putnam, Lakoff, Johnson, and Jackendoff all say things that have at least a superficially similar sound, but as far as I can tell do not have in mind the same sort of technical directions I am suggesting here.

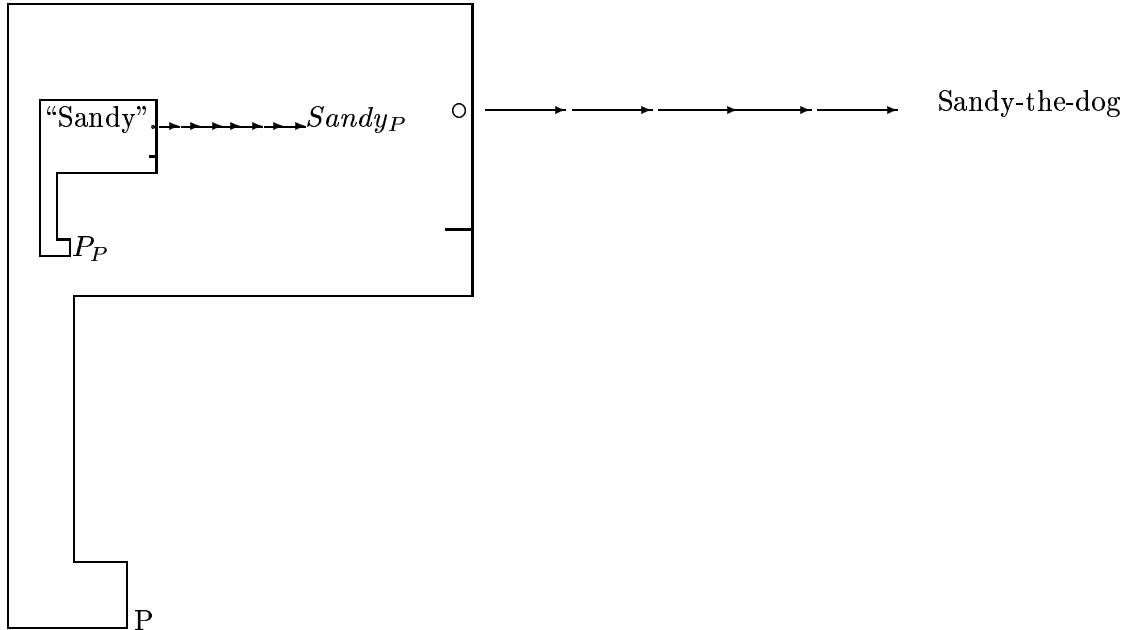


Figure 1.4: Getting outside oneself.

neural wiring) to highly detailed parts of the rest of the body. Thus there is a “motor cortex” that involves a collection of neurons that are linked (via motor neurons) to bodily muscles; and a “sensory cortex” similarly linked to sense organs. These cortices are furthermore laid out spatially in the brain in rough correspondence to the shape of the corresponding distal body parts. In addition, there is another homunculus, in the midbrain (the tectum, to be precise) where various sensory modalities, including proprioceptive information, is combined into a map of the current body position and immediate surrounds insofar as it is known to the individual. Thus the notion of an inner self is no so far fetched. However, this is a far cry from isolating precise structures that play the roles I am outlining in this paper.

1.5 Beyond the body

Now, the hard part. How do we go beyond direct-wiring (body parts) to other pieces of the world? How do we get out to our dog Sandy? Here I propose that we use geometry. Our mental space has a foot(-unculus) in it. Let us use that to mentally step off several feet in our mental space until we get to $Sandy_P$. The result is the number of feet (so to speak) that we take $Sandy_P$ to be in front of us. Of course, all this is in our mental space, not out there in the real external space between us and Sandy. But there is an obvious extension to the external space: the actual number of actual foot-lengths between us and the actual Sandy. We can even get at this, by physically laying out rulers (or shoes, or our own foot again and again) until we bump into the dog (actually and also seen/felt in our mental space). We might then find that our guess was right, or off the mark.

See Figure 1.4.

Here “Sandy” is the word/sound/token of the dog’s name inside P’s mind. *Sandy_P* is the dog Sandy as far as P is concerned; it is P’s mental representation of Sandy (visual and/or whatever other forms it may involve). Note that “Sandy”—but not *Sandy_P*—is represented by P as being *in* P’s mind/head (i.e., in her self-homunculus). This reflects a further property of internal reference: P takes herself to be a meaner, an entity where symbols are maintained, typically distant from their referents. P does not (usually) regard *Sandy_P* as being in her herself—after all, this is (to her) a dog! But P’s uses of the dog’s name are (for her) events initiated in her head. On occasion, to be sure P might consider *Sandy_P* to be in her own head; for instance, if someone calls this to her attention (“it’s really just a visual image”) or if something makes P suspect that she is deceived about what is really out there. This brings up the issue of error and error-correction, which we address elsewhere in this paper.

We can measure, in our mind’s eye, how many of our footlengths would appear to be required to step off the distance between us (i.e., our inner self image) and Sandy (i.e., our inner *Sandy_P*). This may or may not correspond to the actual number of actual footlengths to the actual Sandy. In Figure 1.4, we have indicated an inner measure of seven footlengths between the inner self *P_P* and the inner *Sandy_P*; and five real footlengths between P and Sandy-the-dog. It appears to P that the dog is seven feet away, but she is deceived, and if she steps out footlengths she will bump into Sandy after only five steps. P’s belief that “Sandy is seven feet in front of me” is false. If this happens a lot to P, she probably will learn to recalibrate the way she “steps off” footlengths in her mind’s eye. This can be thought of, I suggest, as a change in (part of) her notion of footlength. Thus she “re-calibrates” her iref so that it is a better fit with reality.

Thus we use an inner metric (footuncular lengths, say) which has a canonical map to a real physical measure (our real footlength). The latter can be checked by carrying out a physical measurement (stepping off so many footlengths to Sandy, and thus we can determine whether our estimate was accurate. The meaning, then, of “that thing three of my footlengths in front of me” is whatever may happen to be three of my actual footlengths in front of me. This depends on our knowing that “my foot” means my foot, but this is determined by links in my head (iref) between that phrase and my footunculus. For the footunculus *is* connected to my actual foot; and the iref *is*, we may suppose, largely formed from contacts with my socio-linguistic community: I learn to link the sound “foot” to my footunculus.

We build our inner world, and carry on referring-behavior with respect to it. Often this behavior fits well with the outer world around us, in the sense that we get the expected outcomes. When we do not, we invoke calibration behavior to reset our world view; that is, we create a larger world in which our old world is a datum to which “didn’t fit” is applied. Thus we can distinguish things as imaginings or as real within my world. I want to sit down, so I reach for the chair behind me, in my world. My real arm also moves behind the real me, and if the latter does not meet a real chair, I will (typically) not get the expected hand-meets-chair result in my world, so I recalibrate my world by taking the chair behind me away. That is, I stop believing there is a chair close behind me, but I continue to keep the chair-close-behind-me idea, now marked as false (poor fit).

However, the above does not really offer any insight into *how* the inner measuring is done. I have no detailed theory to offer, but perhaps some further comment is useful.

To think one is thinking, is to attend to one's attending, that is, to have an object in mental space that plays the role of—stands for—one's attending mechanism. There is then a natural covariance between the mental-space attender (one's inner self-concept) and the actual attending mechanism. This is the beginning of a map between inner and outer worlds. Next we need an inner metric, and a way to calibrate it with external inputs. (Actually, for mere comparison to another's mind, we might not need calibration, this is unclear to me.) But the inner metric is given in the inner world; it comes with a measuring device, with ground zero being the model-attender. I suggest that time of scan may provide a basis for the metric. To scan from the (inner) attender out through mental space to (inner) object q , takes time t , and then t is then a sort of distance d that q is from the attender. Now this is fraught with problems, for I can think about (attend to) stars that are light-years away (even in my conception of them) in as little time as I can think about my home one mile away. So I am tempted to propose a basic contiguous-scan device that moves out through mental space without taking shortcuts. But this is in some sense irrelevant since it is all adjustable according to recalibration as hinted at above and further described below.

Now how do we link up inner and outer units? Perhaps we use standard units, such as one's own body size. My mental foot has an inner-measured size, and the outer foot has a real length; I move my foot according to my mental measures, and re-calibrate if things go wrong. What makes something my mental foot? Simply that it moves when and only when my real foot moves.⁸

If I am totally disconnected from bodily parts that I can move and sense, then it may be that I cannot refer to objects in the outer world. I will have no basis for measuring, so ability to judge that this is close or far, that this is up or down, large or small, in front or behind, no way to calibrate. But I suspect that even more damage than this is required to totally break the map.

How is calibration done? Once an expectation as to what a given portion of the inner world will do, its failure to do so occasions an effort to revise some portion of one's beliefs (i.e., of one's inner world) so as to fit (and explain) what was observed. There is no one right way, of course. Things like Occam's Razor will probably be important. And now we can see that we are in effect simply being everyday scientists, revising our theories when data so dictates. If I do not get a hand-on-chair observation, I revise my belief that the chair is within arm's length of me. I move it from arm's length to a further distance. And so doing is a calibration.

How does this help with intentionality? How does it let us specify when I mean the same chair as you? If we both have thoughts involving portions of our worlds, respectively, that are arm's length behind me, then we are thinking of whatever occupies the same piece of space, assuming our arm's length notions are calibrated. We may, for instance, both have been present when my arm was measured for a shirt, and may both have heard that it is 32 inches long. We may also each have an inch notion, calibrated again and again by us as children through the use of rulers in art class.

It would be highly instructive to set up a simple, one-dimensional world, and work out

⁸In the normal course of things. If I am an amputee, I may have foot-motion sensations, but there will be a tell-tale lack of regularity, in that I will not have control over the foot. I cannot make my inner foot move at will anymore because the motor nerves do not reach muscles. In this case I will have evidence that my inner foot does not correspond to a fully functioning external foot.

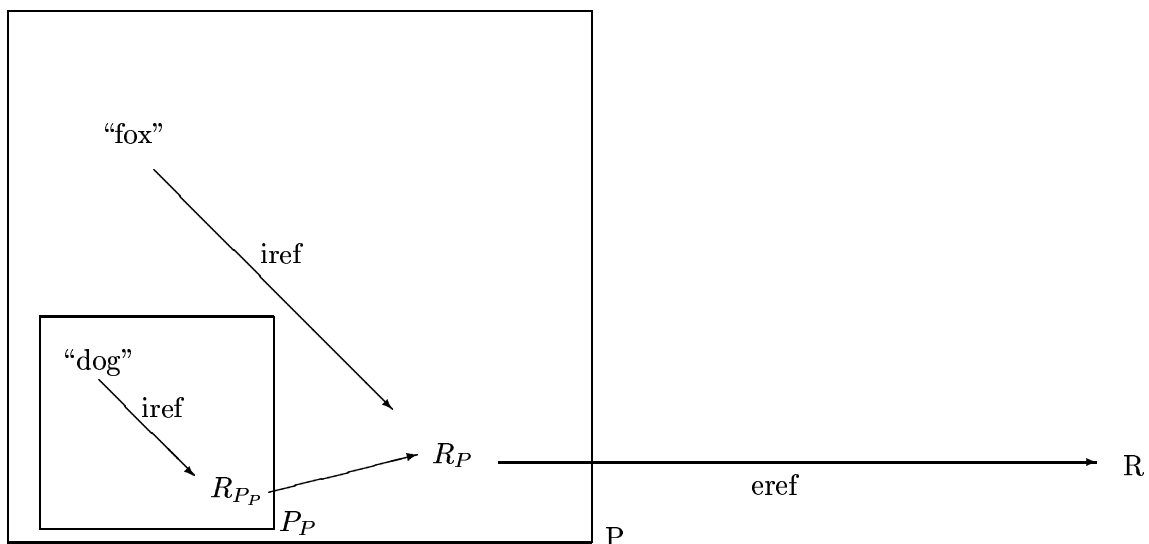


Figure 1.5: Error Recovery.

the details for a simple organism that can think and calibrate.

1.6 Error and recovery

We return briefly to the issue of error, since it is a key one in the literature on intentionality, and in particular a central theme of our *Part I*. We consider the example of seeing something, thinking it is a dog, and then deciding it really is a fox but appeared at first to be a dog. In terms of our pictorial models, a complex picture is formed, as indicated in Figure 1.5. The idea is that P first thought “there is a dog” but then (the state as pictured) “I thought that was a dog, but it’s a fox.”

Thus there is a fundamental belief-layer, in which the current view is taken at face value, and only later can be called into question by means of passage to a new (perhaps more introspective) view (which then in turn is taken at face value). Whatever the current view is, is believed and indeed taken as being reality.⁹ That is, instead of there being a “true” box which marks those sentences in one’s head that are judged true, there is rather a “doubt” box which marks those that have come to be doubted. The default is to believe. It is a reflective act that shows a belief to be, after all, merely something in one’s head. Only at this point is the belief separated—in the reasoner’s mind—from that which is it about. Up to that point there is no distinction made.¹⁰

Default reasoning is a hot topic in artificial intelligence. And it is largely about drawing defeasible conclusions: ones that might be in error, and that additional information might cause us to retract. Thus the present proposal can also be construed as a suggestion toward

⁹Pictorial representations are not always useful here, and I do not mean to suggest the entire issue of belief and belief-correction is a mere matter of pictures).

¹⁰This has interesting points of agreement with research in cognitive psychology. See for instance Dan Gilbert, *How Mental Systems Believe*, to appear, *Amer Psychologist*. Described in Bruce Bower, *True Believers*, *Science News*, vol 139, Jan 5 1991, pp 14-15. Also Flavin et al (ARD).

a mechanism for default reasoning.

Fodor compares MacBeth reaching for a presumed dagger, and a frog snapping at a moving dot (a fly or a bee-bee). MacBeth, he says, can come to see his error, but the frog cannot be said in any serious sense to be able to error: it is built so that it snaps at moving dots, and presumably has no inner distinction to make about something failing to be a fly (or even being found not to be a moving dot after all once snapped at). According to Fodor, the frog merely has moving-dot-appearance detectors, whereas MacBeth has real dagger detectors.

Despite Fodor, we also merely have dagger-appearance detectors; that's all we ever detect: appearances. But we have other processes, that can assess whether a given appearance-detector's firing is veridical with respect to other criteria (e.g., does it cut, can it be grasped, do others see it) which are also appearance detection. And we also have mechanisms for trying to sort out things when confusion (non-veridicality) sets in. This is how "though he [MacBeth] and I both make mistakes, we are both in a position to recover," as Fodor says (TOC p. 107). Recovery is a key notion, although Fodor does not dwell on it. He is less interested in looking inwardly at our behaviors than outward toward the distal stimuli. But recovery is an inner comparison between what is and what isn't, in the agent's world; that is, between what the agent takes to be really out there, and what she takes to be in her mind. To do this she needs to have both of these represented in her mind, i.e., she needs a robust world-model that includes a self-model. The self-model is where the not-really-out-there things go: they are one's thoughts, feelings, etc. And if we are to attribute our mistakes to them, then we need them to be active parts of us.

Let us consider error for kind terms. Let us use Fodor's example of "horse"—not just my horse Henry, but the general notion of a horse. In the view I am advancing here, I have a complex description that constitutes my meaning for the term "horse". Part of it is explicit (non vicarious): animal, typically 4-legged, looks more or less like *this*. And part of it is vicarious: the biological kind specified by technical features of genetics, evolution, and other things experts know but I do not.

Now I see an object in front of me in dim light. I cannot verify all parts of my "horse" description, not even all the explicit parts. But I verify enough of them to decide it is a horse. I am guessing, but not wildly.

Now there are several ways error might have occurred here:

1. error of public/expert use: I may discover others do not use "horse" the way I do. This is the case of Joe's arthritis and Sue's Purdue.
2. error of term picking: I may have slipped up in manipulating my tags and taggees, and picked the wrong tag, having gotten "horse" instead of "donkey". I would then claim to have meant the right thing (donkey) but grabbed the wrong word ("horse"). I had the right R_P .
3. error of identification: The lights may brighten and I may now see that there are stripes on the animal: it is a zebra. I decide that what I *thought* was a horse is in fact a zebra. I did not merely pick the wrong word, I had the wrong R_P .¹¹

¹¹This seems to be the kind of error Fodor (1987, p. 99ff) is concerned with. I hope that this provides

At least one issue is left unclear in this example, however. For the tag used for a taggee is itself, at least sometimes, part of (or even all of) what we know as the description of the taggee. In that case, we cannot get the tag wrong but the taggee right (case 2 above). Putnam’s elms come to mind. If all we know about elms is that they are something-or-other tagged by the term “elms”, then we cannot get the tag wrong without also getting the taggee wrong. This seems to require some modification of my earlier claim that we have no control over R_P : in a case like this we do have some control. In fact, a case like this is largely one of pure formalism (if we really know *nothing at all* about elms, not even that they are physical objects). The entire meaning is then in terms of a tag applied to something created only for the purpose of attaching that tag to it! So we create both tag and taggee here. This will come up again in the final section on reflection.

This also points up that I have been quite vague as to what, in general constitutes a taggee or object in mental space.¹² I cannot offer any better analysis at this time, however.

I am similarly vague about where descriptions fit into the S-to- R_P picture of iref. I am not sure there is a definite answer here: perhaps descriptions can play three roles: as S (name), as R_P (named) and as part of what cements S to R_P . One problem with giving a straightforward explanation of this is that the entire mechanism is dynamic, as is brought out more forcefully in my discussion of error and especially of reflection in the final section.

1.7 Functional role semantics vs. iref

The above view of meaning suggests some alterations in functionalist treatments of reference. For instance, it suggests that we can have reference or content to thoughts *without* external embedding other than the space-time features specified. We need a temporally manipulable space of tags and taggees, but any computer has this. Inputs and outputs external to the system are not essential, iref may still be present. Notice the word “may”. There is no claim here—indeed a denial—that every computer (programmed and running) has iref. But the presence of iref depends not at all on events outside the system. It depends only on the right sort of set-up *inside* the system.¹³

This opens up the possibility that two meaners could be thinking the same thing (about a game of chess, say) with no contact whatsoever with the outer world. The chess-game vs. war-game scenario of Georges Rey does not apply here, since it is not multiple external maps that determine reference (chess vs. war) but internal tagging structure, and it may happen that these are the very same in the two meaners. What would make the activities chess games rather than war games would be details about the tag network, such as the feature that there is a *general* rule being applied as to how a knight always moves, for instance, rather than a one-time rule that says this knight moves in this way only in this particular game. Of course, truly distinguishing between war and chess in a thoroughly

a solution to the problem. Fodor’s own attempt seems to me not to work, since it appears to remain fully verificationist with no means of appeal to unavailable or vicarious information.

¹²Even a tag is an object in mental space, since we can turn our attention to the tags, thereby tagging (quoting, perhaps) them with yet further tags. This is important in recognizing and dealing with error, in more generally in considering alternative possibilities (thinking).

¹³See Rapaport for a very similar suggestion.

convincing way involves a myriad of very human details that we cannot begin to analyze in the present state of theory.¹⁴

Do the players in the “game” think of it as war, or chess? Only they can tell! It’s their world. But can there be an intrinsic meaning? Can there even be a sense in which an outsider (outside their vat, so to speak, to relate this to the new Putnam) can make sense of the notion that inside that vat the beings are thinking of war (or chess)? There’s a strong intuition that even if it should turn out that all my thoughts and memories are dreams (or ruminations) and there is no outer world at all, still I have chess and war ideas and so could someone else, and my chess and her chess would both be chess, i.e., the same in crucial respects, one would not be chess and the other war. Indeed, the geometric reference idea seems to let us both match unicorn concepts, for that matter. We do not need an external world to appeal to here.

Another modification iref suggests is simply itself: another dimension should be adjoined to the nature of mental states. It is not enough to covary with an external entity and be the object of inferential manipulations. For this can be done by a one-tiered system, which has no internal meaning (of its own) and hence no error (of its own). Iref buys us error (and with it the chance of undoing error) and also buys us ourselves. I.e., the tag-mechanism allows the possibility of constructing a division in the mental space into two parts: the tags and the taggees; when error pushes us to revise the tagging network, we compare the old (“errorful”) status and take it to be false, not true of the world, not “real”. But if not real, then what is it? It is that part of the world that represents (sometimes incorrectly) other parts: it is a meaner (us!). In seeing how our ideas differ from reality we see a thing (ourselves) as approximating but different from (part of but not the whole) world.

As a consequence, two *functionally* equivalent engines may have distinct mental states: indeed one may be one-tiered and have no mental state at all (on the view I am advocating here) and the other may be two-tiered and have rich mental states.

Finally, the internal semantics urged here (iref) is no mere stance (*a la* Dennett). It is a real design feature of the system in question.¹⁵

1.8 Reflection and geometric semantics: a summary

We have summarized our proposal in a formula:

$$\text{content}_P(S) = \text{eref}_P(\text{iref}_P(S)) = \text{eref}_P(R_P) = R$$

Roughly, we can get outside to the real world by taking our internal measured universe and mapping it outside through our (real) feet!¹⁶

¹⁴One might imagine a war in which certain tanks are constrained to move by their very nature just the way knights do in chess, and so on throughout the story. But then at some point such a war is perhaps truly a game of chess too. Then the example seems to break down altogether, so we will leave it here.

¹⁵This suggests answers as well to Block’s *Chinese Nation* and Searle’s *Chinese Room* puzzles: if there is iref in the computations then there is meaning/mind/intentionality. It also suggests a way to determine whether two meaners have the same internal meanings for a term: if they employ the same descriptions/tagging networks for it (modulo indexicals). This is not holistic, and allows for *degrees* of similarity of meaning.

¹⁶I hope the reader understands that arms and other body parts do just as well, and that I am not claiming

A simpler case than that of Sandy-the-dog is basic geometry, e.g., a rectangle. What makes a bit of my brain process amount to my thinking of a rectangle? Surely I could do this even if I had never seen or heard of a rectangle. Yet how could my brain-chemistry then have anything to do with things quite different from it? For there are no rectangles in my brain, or at least none of roughly human size, which—let us suppose—is how I imagine my rectangle to be. Here is my answer: we use geometry, to measure our brain-chemical process, in terms of other such standard processes, e.g., the my-foot-idea which is also a brain process.

One might ask how the latter brain-chemical process can be justifiably called a foot-idea? Doesn't that just beg the question? No, for there is an important difference: my real foot is wired to the my-foot-idea in such a way that they covary extremely well. And now, from the foot measuring stick, we can reconstruct the universe as we see it: indeed that is what it is for us to have a view of the world!¹⁷

To get content out there, we first need to get it in here. My immediate meaning for “John” is the John in my world, which is something in my head. That John in my head is related to the real John out there; but how? By the real John's being situated in the real world pretty much the way my world's John is situated in my world. How do we even begin to compare the real world and my world, to see if there are similar situatednesses? By the canonical pinning point: me. My world's right foot maps canonically to my real right foot, as we saw before. My body is highly mapped this way, to homunculi in my brain, a very sophisticated and robust system of covariances. My “John” thoughts, however, do not covary nearly as well with the real John's presence, and this is where distal attempts founder. By going right out there to distal objects the robust proximal maps are ignored and with them the basis for a robust geometry that can in fact capture the distal objects.

We must avoid the a verbal difficulty; when I speak of the John in my world, I mean as an outsider would describe my world. To me of course that simply is John, except when I grow philosophical and try to catch my thoughts as objects. Then I do make a distinction between the real John and the my-world John, but I do so only by going deeper inside, not by going out to the real John. What I do is “reflect”: I create another layer in my world, that has a model of me in it and the (as I would then style it) my-world John sitting in the my my-world head while the real John (as I would style it) sits outside that head. This is probably the sort of thing that drove Kant (and now Putnam) to their dismal conclusion that we can never know or even think about things-in-themselves. But this is not so bad: the whole of science shows how we can make robust progress despite the possibility of total confusion. We get evidence that our inner worlds match pretty well with the real world; not proof, but impressive evidence.

We need to say more about something referred to here and there in the foregoing discussion, namely what constitutes thinking about something. Merely to tag R_P with S is not to think about R_P , at least not in the sense of thinking aimed at getting a better understanding of R_P . On our view, getting a better understanding amounts to recognizing potential errors or gaps in R_P (as the thinker's view of some hypothesized external R) and

that ideas held by footless persons thereby cannot have external significance! I discuss this issue more in the earlier section on body-reference.

¹⁷How do we account for colors and other qualia? This is a terrifically hard one we shall not address further here.

trying to bring R_P more in line with the real R. But how can this be? It seems to fly in the face of all we have said here: the thinker P is constrained to do all her mental activity only with the materials at hand, namely things in her mind or mental space. There is no R there.

This is where reflection comes in. Recall that error (in one form) amounts to finding a discrepancy in one's own (delayed-verification of one's) descriptions. What is the thinker to do on discovering such? The answer that I propose is as follows: she creates a new item in her mental space, that is (to her, now) the real thing out there—this was what R_P *was* to her formerly but now she decides her R_P was a mistaken identification of reality, merely how reality (which she now views her *new* item as being) *appeared* to her through her confused conceptual/perceptual mechanisms (her glasses were dirty, the light was bad, whatever).

She has thus “reflected” the entire tagging apparatus into an object for scrutiny. She has not really gotten “out there” to the real R in itself (as Kant warned cannot be done). But she may have gotten closer to it (a better approximation to it).¹⁸

Perhaps then, in a sense, we never do refer to real external things. Perhaps we can yield on this point to the “renegade Putnam” (as Devitt and Sterelny style Putnam’s thinking since around 1980). But also perhaps we do not need such reference, if iref can through reflection and space-time-situated error lead us to better and better world models. Indeed, our obvious success in dealing with (surviving in) the world speaks strongly for exactly this.

Thinking, then, I suggest amounts to considering alternatives, e.g., weighing them as to their more plausible “reality” (as seen in our mental space). Performing this involves frequent reflections (in-turnings) of taggee into seen-as-taggee, i.e., a further flexing of our control of symbolism.

This notion of meaner-control of meaning often appears in the literature, but in the form of asides or casual remarks (e.g., Fodor’s LOT, p. 191). Here I have argued that this is one of the primary features of mind and meaning, and provides us with the luxury of making mistakes and thereby improving our world pictures. One curious feature of the view urged here is that there is no reference without self-reference. The meaner herself is an essential part of her own mental space, for she takes the measure of reality (in her own inner terms) by means of her own situatedness not only as a physical object but as a thinker to whom things can appear different from what they are.

In sum, we have progressed through three different pictures of meaning, from a 1-tiered model (which is the basic distal view) to the two-tiered, to the internal body-in-the-world-model where we can recapture external (distal) reference via internal geometry and where active representation is possible as well as reflection/thought. I suggest that this line of investigation may lead to a characterization of what thinking is. Much of what I have proposed may seem obvious to a neuroscientist; certainly the tectal maps are well-known. But the reflection capability is, as far as I know, new, and apparently nothing has as yet been discovered in the brain that may play such a role.

¹⁸Dwelling on such matters leads one to the frustrations that Berkeley *et al* suffered, and can also bring on bouts of solipsism and/or despair. But thinkers of past centuries tended to ignore the positive feature of better and better approximations. Perhaps our age of frequent scientific advances makes us more aware of this possibility.

1.9 Virtual reality machines

Recently there have been studies involving so-called “virtual reality machines.” See, for instance, *Sci. Amer.* Feb 1991, vol 264, no 2, p. 134, “Reach out” by Gary Stix. These machines present a user with sensory input very much like that in a normal sensory setting in which an external object is “sensed” via the usual signals (optical and otherwise). In the case of these machines, however, there is no external object (at least of the usual sort). Instead the machine simulates the sensory inputs that would normally be produced by the external object; thus visual and other inputs are manufactured in the machine and presented to the user. The effect can be quite startling. early versions of such machines include the widely used aircraft simulators pilots use to gain experience in take-offs and landings. The newer machines are more sophisticated and have applications far beyond piloting aircraft.

What is of interest for us, is that such machines provide a setting in which, in a sense, *erref* fails but *iref* succeeds. External objects appear to be present, but they are not. We might phrase this by saying that the objects are there in the subject's *view*. Yet, this is always the case. A rainbow is, in some significant sense, not there at all in the world. There is in fact no particular physical entity situated just where there appears to be a rainbow. The appearance results from our limited visual field and the angles of refraction of light on water droplets. There is of course something there, but the something is not very much like what we “see”. What is actually there are water droplets and refracted light going in rather subtle directions resulting in an ordered array of wavelengths being perceived at any small region where a viewer stands. It is, in an important sense, an illusion. And yet, we can point to it, discuss it, analyze it.

Similarly with a virtual reality machine. Two users can agree (or disagree) on details of the perceived world. Their words are, then, surely about something. But the something is complicated. We can even say the same about a dream, even though this is not a shared access to perceptual data; a dream has features that can be discussed. Finally, even in the most ordinary of settings, e.g., seeing an actual table right in front of oneself, there is here too a view that is best understood in terms of the viewer. After all, we now know that a (wooden) table is not a totally solid homogeneous mass; rather it is mostly water in countless cells that in turn are jostling atoms, and so on. Our everyday sense of it is perfectly adequate (even, perhaps, excellent) for everyday purposes, but then so is our sense of a rainbow, and our sense of the virtual-reality experiences. Only when we need to look deeper do we find it useful to relinquish these views for others. And we most likely have not yet discovered a “final” level that is the “one true” reality. Yet we surely should not conclude then that we have not yet ever experienced reference in our thinking and talking!

What is, I think, essential for there to be reference, is a reliable recourse to data. As long as we can reliably point to color data (i.e., take another look at the colors), we can speak meaningfully of a rainbow, and know what we mean. Experiential tableness is there, whether or not we wish to argue that a solid undifferentiated mass is there that will stand up to all investigation.

Now, this is not enough. For we do want to be able to distinguish between experiences we regard as real, and ones that are illusory. However, we never get a final answer on this. Rather we go by the evidence. As long as the evidence appears consistent with some interpretation, we may feel comfortable with it and go on using it. But when the evidence

(in the form of more data) reaches a certain discomfort level, we start to change our view. Thus the case of a missing limb: at first we may refuse to believe the limb is not there, for we may still “feel” it there. But eventually experience starts to erode that view, as we re-calibrate our inner metric: the proprioceptive foot-length no longer is a good measure for us, so we rely more on other measures.

Bibliography

1. Barcan Marcus, R. (1983) Rationality and believing the impossible. *J. Phil.* 80:321-338.
2. Brentano, F. [1874] *Psychologie vom empirischen Standpunkte*. Leipzig.
3. Churchland, P. [1984] *Matter and Consciousness*. MIT Press.
4. Cummins, R. [1989] Methodological reflections on belief. Paper presented at the 1989 meeting of the SPP. Tucson, Arizona.
5. Dennett, D. (1978) *Brainstorms*. Montgomery, VT: Bradford Books.
6. Dennett, D. (1987) *The Intentional Stance*. MIT Press.
7. Dennett, D. (1988) Granny's campaign for safe science.
8. Derrida, J.
9. Devitt, M. [1981] *Designation*. Columbia University Press.
10. Devitt, M. [1984] *Realism and Truth*. Blackwell.
11. Devitt, M. [1984] Thoughts and their ascription. In *Midwest Studies in Philosophy*, vol. 10; P. A. French, T. E. Uehling, and H. K. Wettstein, editors. University of Minnesota Press.
12. Devitt, M. [1988] Realism without truth: a response to Bertololet. *Analysis* vol. 48, pp. 198-203.
13. Devitt, M. [1989] A narrow representational theory of mind. In *Representation: Readings in the Philosophy of Mental Representation*; S. Silvers, editor. Kluwer.
14. Devitt, M. [forthcoming] Against direct reference. In *Midwest Studies in Philosophy*, vol. 14.
15. Devitt, M. (1990) Naturalistic representation. Manuscript.
16. Devitt, M. and Sterelny, K. (1987) *Language and Reality*. MIT Press.
17. Dretske, F. (1988) *Explaining Behavior*. MIT Press.
18. Elgin, C. (1985) Translucent belief. *J. Phil.* 82:74-91.
19. Fodor, J. (1979) *The Language of Thought*. Harvard University Press.
20. Fodor, J. [1981] *Representations*. MIT Press.
21. Fodor, J. [1987] *Psychosemantics*. MIT Press.
22. Fodor, J. [1990] *A Theory of Content*. MIT Press.
23. Kihlstrom, J. [1989] Implicit cognition and the cognitive unconscious. Paper presented at the 1989 meeting of the SPP. Tucson, Arizona.

24. Kripke, S. (1979) A puzzle about belief. In: *Meaning and Use*, ed. A. Margalit, pp. 234-283. Holland: Dordrecht.
25. Kripke, S. (1980) *Naming and Necessity*. Harvard University Press.
26. Locke, J.
27. Lycan, W. [1987] *Consciousness*. MIT Press.
28. McCarthy, J. (1979) Ascribing mental qualities to machines. In: *Philosophical Perspectives in Artificial Intelligence*, ed. M. Ringle, Humanities Press.
29. McCarthy, J. [1979] First order theories of individual concepts and propositions, *Machine Intelligence* 9, 129-147.
30. Minsky, M. (1968) Matter, Mind, and Models. In: *Semantic Information Processing*, ed. M. Minsky, pp. 425-432. Cambridge: MIT Press.
31. Nagel, T. What is it like to be a bat?
32. Odell, S.J. (1984) On the possibility of natural language processing: some philosophical objections. *Theoretical Linguistics* 11:127-146.
33. Perlis, D. [1981] *Language, computation, and reality*. Ph.D. thesis, University of Rochester. Tech Report TR95.
34. Perlis, D. [1986a] On the consistency of commonsense reasoning, *Computational Intelligence* 2, pp. 180-190. Reprinted in M. Ginsberg (ed.) *Non-Monotonic Reasoning*, Morgan Kauffman, 1987, pp. 56-66.
35. Perlis, D (1986b) What is and what isn't. Symposium on intentionality, Society for Philosophy and Psychology, Johns Hopkins University.
36. Perlis, D. [1987] How can a program mean? Proceedings, International Joint Conference on Artificial Intelligence, August, 1987, Milan, Italy.
37. Perlis, D. [1988] Commonsense set theory. In *Meta-Level Architectures and Architectures*, P. Maes and D. Nardi (eds.) North-Holland.
38. Perlis, D. (forthcoming) Intentionality and defaults. *Advances in Human and Machine Cognition*, K. Ford and P. Hayes (eds.). JAI Press.
39. Perlis, D. and Hall, R. (1986) Intentionality as internality. *Behavioral and Brain Sciences*, 9(1), pp. 151-152.
40. Popper, K. (1965) *Conjectures and Refutations*. Basic Books.
41. Putnam, H. (1970) Is semantics possible? In H. Kiefer and M. Munitz (eds.) *Language, Belief and Metaphysics*, SUNY Press.
42. Putnam, H. (1975) The meaning of 'meaning'. In K. Gunderson (ed.) *Language, Mind and Knowledge*, Univ of Minnesota.
43. Putnam, H. [1975] *Mind, Language and Reality: Philosophical Papers*, vol. 2. Cambridge Univ Press.

44. Putnam, H. (1981) Reason, truth and history. Cambridge U. Press.
45. Putnam, H. (1987) Representation and reality. MIT Press.
46. Pylyshyn, Z. (1984) Computation and Cognition. MIT Press.
47. Quine, W. [1953] From a Logical Point of View. Harvard Univ Press.
48. Rapaport, W. [1976] Intentionality and the structure of existence. Ph.D. thesis, Indiana University.
49. Rapaport, W. [1988] Syntactic semantics: foundations of computational natural-language understanding. In J. Fetzer (ed.) Aspects of Artificial Intelligence, Kluwer, pp. 81-131.
50. Rock, I. The logic of perception.
51. Rosenschein, S. and Kaelbling, L. [1987] The synthesis of digital machines with provable epistemic properties. SRI Tech Report 412.
52. Sayre, K. (1986) Intentionality and information processing... Behavioral and Brain Sciences, 9(1), pp. 121-138.
53. Searle, J. (1980) Minds, brains, and programs. Behavioral and Brain Sciences 3, 417-424.
54. Sloman, A. (1986) Reference without causal links, Proceedings, 7th ECAI, July 21-25, 1986, Brighton, UK. 369-381.
55. Steels, L. (1986) The explicit representation of meaning. Proceedings, Workshop on Meta-Level Architectures and Reflection, Sardinia, October 1986.
56. Stich, S. (1984) From folk psychology to cognitive science: the case against belief. Cambridge: MIT Press.
57. Suchman, L. [1985] Plans and situated actions, Tech Report ISL-6, Xerox Palo Alto Research Center, February 1985.
58. Tarski, A. [1936] Der Wahrheitsbegriff in den formalisierten Sprachen, *Studia Philos.*, 1, 261-405.
59. Thagard, P. (1986) Parallel computation and the mind-body problem. *Cognitive Science*, 10, 301-318.
60. Waltz, D. and Boggess, L. (1979) Visual analog representations for natural language understanding. Draft, Advanced Automation Group, Univ. of Illinois.