Representation and Indication

Robert Cummins

Pierre Poirier

Department of Philosophy

University of California at Davis

rcummins@ucdavis.edu

1. Two kinds of mental content

This paper is about two kinds of mental content and how they are related. We are going to call them representation and indication. We will begin with a rough characterization of each. The differences, and why they matter, will, hopefully, become clearer as the paper proceeds.

1.1 Representation

Some authors (e.g., Schiffer, 1987) use 'mental representation' to mean any mental state or process that has a semantic content. On this usage, a belief that the Normans invaded England in 1066 counts as a mental representation, as does the desire to be rich. This is not how we use the term. As we use the term, a mental representation is an element in a scheme of semantically individuated types whose tokens are manipulated — structurally transformed — by (perhaps computational) mental processes. The scheme might be language-like, as the language of thought hypothesis asserts (Fodor, 1975), or it might consist of (activation) vectors in a multidimensional vector space as connectionists suppose (e.g., Churchland, 1995). Or it might be something quite different: a system of holograms, or images, for example.

On one popular theory of the propositional attitudes, having the belief that the Normans invaded England in 1066 involves tokening a mental representation with the content that the Normans invaded England in 1066 — writing in the Belief Box a representation that means that the Normans invaded England in 1066.[1] That theory takes the content of the belief to be the same as the content of the implicated representation, but distinguishes representations from attitudes, taking representations, as we do, to be expressions in a scheme of representational types. We think this theory of the attitudes is seriously flawed (see Cummins, 1996), but one thing it surely gets right is that representations should be distinguished from attitudes.[2]

1.2 Indication

It is useful to begin with some influential examples.

- <u>Thermostats</u>: The shape of the bimetallic element (or the length of a column of mercury) is said to indicate the ambient temperature.
- <u>Edge detectors</u>: Cells in the primary visual cortex (V1) strongly respond to edges in the visual field, that is, linear boundaries between light and dark regions.[3]
- "<u>Idiot lights</u>". Most cars have indicator lights that come on when, e.g., the fuel level is low, or the oil pressure is low, or the engine coolant is too hot.

'Indication' is just a semantic-sounding word for detection. We are going to need a way to mark the distinction between the mechanism that does the detection and the state or process that is the signal that the target has been detected. We will say that the cells studied by Hubel and Weisel (1962) are indicators, and that the patterns of electrical spikes they emit when they fire are indicator signals. Rather less obviously, a mercury thermometer is an indicator, and the length of the mercury column is the indicator signal. Similarly, the bimetallic element found in most thermostats is an indicator, and its shape is the signal. [4]

2. Indication and representation contrasted

It is commonplace to think of indication as a species of representation. Indeed, one very popular theory of representational content has it that, at bottom, representation just is, or is inherited from, indicator content.[5] We think there are some good reasons to keep the two distinct.

<u>Indication is transitive, representation is not</u>. If S3 indicates S2, and S2 indicates S1, then S3 indicates S1. Aim a photosensitive cell at the oil pressure indicator light in your car. Attach this to a relay that activates an audio device that plays a recording of the sentence, “Your oil pressure is low.” If the light going on indicates low oil pressure, so does the recording. Indeed, there is already a chain of this sort connecting the pressure sensor and the light. Representation, on the other hand, is not transitive. A representation of the pixel structure of a digitized picture of my aunt Tilly is not a representation of my aunt Tilly’s visual appearance, though, of course, it is possible to recover the later from the former. To anticipate some terminology we will use later, a representation of the pixel structure is an <u>encoding</u> of my aunt Tilly’s visual appearance.[6]

<u>Indicator signals are arbitrary in a way that representations are not</u>. This actually follows from the transitivity of indication. Given transitivity, anything can be made to indicate anything else (if it can be detected at all), given enough ingenuity and resources. (This is what makes it tempting to think of words as indicators: they mean

something, but they are arbitrary.) It follows from the arbitrariness of indicator signals that disciplined structural transformations of them are not going to systematically alter their meanings. Susceptibility to such transformations, however, is precisely how representations earn their keep. Consider, for example, a software package that takes a digitized image of a face as input and "ages" it, i.e., returns an image of that face as it is likely to look after some specified lapse of time. Nothing like this could possibly work on an input that was required only to indicate a certain face, because there is no correlation between the physical characteristics something must have to be a signal that indicates the appearance of my face at age 18 and the physical characteristics of my face at age 18. You could design a device that <u>seemed</u> to do this, of course. Given a name and a target age, it would retrieve from memory its most recent picture of the person named (if it has one), and age it by the number of years equal to the target date minus the date the picture was taken. But surely(?), respectable cognitive scientists would not be fooled by such a device. They will know that you need a representation, not just an indicator signal, to get the job done, and they will infer the internal representation.[7] Similarly, while it could conceivably be useful to detect the utterance of a particular sentence ("Look out!"), there is no possibility of using a mere indicator signal to effect an active-to-passive transformation. A representation of the sentence's phrase structure, on the other hand, actually has, by hypothesis, the relevant structure, and hence formal transformations—transformations that systematically alter structure--of a representation of that phrase structure can easily be constructed to give a representation of the structure of the corresponding passive.

It is, therefore, a consequence of the nature of indication that the structural properties of an indicator signal (if it has any) have no significance.[8] Indicators "say" that their targets are there, but do not "say" anything about what they are like. Representations, on the other hand, mirror the structure of their targets (when they are accurate), and thus their consumers can cognitively process the structure of the target by manipulating the structure of its representation. But representations, unlike indicator signals, are typically silent concerning whether their targets are "present": they are not, except incidentally and coincidently, detector signals.

Indicators are source dependent in a way that representations are not. The cells studied by Hubel and Weisel all generate the same signal when they detect a target. You cannot tell, by looking at the signal itself (the spike train), what has been detected. You have to know which cells generated the signal. This follows from the arbitrariness of indicator signals, and is therefore a general feature of indication: the meaning is all in who shouts, not in what is shouted.[9]

In sum, then, indication is transitive, representation is not. It follows from the transitivity of indication that indicator signals are arbitrary and source dependent in a way in which representations are not, and this disqualifies indicator signals as vehicles for structure dependent cognitive processing. Representation is intransitive, non-arbitrary and portable (not source dependent), and therefore suitable for structural processing. Indicator signals "say" their targets are present, but "say" nothing about them; representations provide structural information about their targets, but do not indicate their presence. Indicator signals say, "My target is here," while representations say, "My target, wherever it is, is structured like so."

3. Portable indicators and the language of thought.

In principle, indicator signals can be distinguished into types so as to reduce source dependency. Rather than label the low fuel and low oil pressure lights, one could make the lights different sizes or shapes or colors. This amounts to building an arbitrary source label into the form of the signal. The result is a portable signal. It is not obvious, however, what advantage there might be to such a system beyond the fact that you could tell in the dark whether it is oil pressure or fuel that is low. To appreciate the value of portability, we have to imagine a more complex system.

Consider, then, the system called LOCKE (see, e.g., Cummins, 1989, p. 37ff).

INSERT FIGURE 1 ABOUT HERE

A TV camera is attached to a card punch device. The pattern of holes punched in a “percept” card depends on the structure of the optical stimulus. The matcher is equipped with a set of “concept” cards. A match is achieved when every hole in a concept card matches a hole in a percept card (barring chads, of course). When a match is achieved, the word printed on the back of the concept card is displayed.

It would be no mean feat to build such a device. A Frog detector, for example, would have to accommodate the fact that frogs come in various postures, colors and sizes, against various backgrounds, with various neighbors, etc. Even color and shape detection will have to accommodate variations in distance, in lighting and in angle of view. For LOCKE to work well, a huge variety of visual scenes whose only common feature is the presence of a frog or red or square will have to produce cards that have the same characteristic sub-pattern. The “analysis” box thus amounts to a sophisticated vision system capable of detecting very “high level” distal properties.

As this example illustrates, the physical characteristics of the signal — punch patterns — can be arbitrarily different from the physical characteristics of the detector itself, and from the physical characteristics of the target. Heat detectors do not have to get hot, nor must motion detectors move. Unlike the sorts of indicators studied by Hubel and Weisel, however, Locke’s indicator signals are type identified by their structural properties (their “form”), and are therefore potentially source independent. This makes it possible to compare concept cards to each other to yield such judgments as that sisters are female, or even that frogs are green, even though none of the corresponding perceptual targets are in the offing. LOCKE, in short, has a crude language of thought, the primitive symbols of which are punch patterns — perhaps single holes — whose contents are fixed by their roles in detection. A punch pattern means <u>square</u>, wherever or whenever it is tokened, if it is the pattern LOCKE’s detection system produces in response to squares.[10] One might then suppose that primitive patterns could be combined in ways that yield complex patterns whose meanings are functions of the meanings of their constituents and their mode of combination. The resulting complex representations will have the semantically primitive indicator signal patterns as constituents, but these will not, of course, function as indicator signals in this new role. Rather, they simply inherit their denotations from their other life as signals indicating the presence of the properties (or whatever) they are now taken to denote.

As things stand, LOCKE cannot represent bindings: It cannot represent the difference between a red square plus a blue triangle on the one hand, and a blue square plus a red triangle on the other. To overcome this limitation, and others like it — to exploit the possibilities of combinatorics — LOCKE will need some syncategorematic representational machinery. The forms that distinguish bindings and the like evidently cannot themselves have their meanings (semantic functions) fixed by their roles in indication. Rather, a complex representation will count as a conjunction or predication or whatever in virtue of how it is processed (Fodor, 1987). The machinery of binding is just whatever form a representation has that causes it to be processed as a binding.

Portable indicator signals, then, provide one way of grounding the capacity to represent in the capacity to indicate (detect). The result is a Language of Thought (LOT) as it is defended by Fodor (1975). There is a characteristic approach to cognition entrained by LOT style representation. It is entrained by the fact that there are basically just three ways that arbitrary mental symbols of the Language of Thought variety can enter into cognitive explanations: As *triggers* for procedures, as *cues* for stored knowledge, and as *constituents* of complex representations. This simply follows from the fact that the structure of a LOT symbol serves only to type identify it. It carries no information about its target, not even the information that its target is present, since, freed from its source, it no longer functions as an indicator signal.

Suppose you are to milk the cow. First you must find the cow. You wander around scanning until your visual system tokens a |cow| — an arbitrary mental symbol that refers to cows. But to visually recognize cows, you need to know how a cow looks. A |cow| contains no information about how cows look, and so it isn't what psychologists, at any rate, would call a visual concept. But knowledge of cows is what you need, for it is knowledge of cows, including tacit knowledge about the sort of retinal projections they tend to produce, that makes it possible for your visual system to token a |cow| when you see one. So the Mentalese |cow| does no work for the object recognition system, it just signals its output, functioning as an indicator signal.

Tokening a |cow|, we may suppose, *triggers* the next step in the plan. Needing to locate the udder, a mental word is totally useless unless it happens to function as a retrieval *cue* for some stored knowledge about cows. Faced with actually having to deal with a cow, the burden therefore shifts again from the symbol to your stored knowledge, because the symbol, being arbitrary, tells you nothing about cows. So it turns out that it isn't because you have a Mentalese term for cows that you find the cow and get the milking done, it is because you have some stored knowledge about cows — some in your visual analysis system, some higher up the cognitive ladder. Mentalese |cow|s could play a role in stored knowledge about cows only as constituents of complex representations — |cows have udders between their back legs|, for example — that are, on the Mentalese story, implicated in the possession of stored knowledge about cows.

LOT stories therefore lead inevitably to the idea that it is really stored knowledge, in the form of systems of LOT “sentences”, that does the explanatory work. It is worth emphasizing that there is a big difference between appealing to the fact that one has a primitive mental symbol referring to cows, and appealing to the fact that one has a lot of knowledge about cows. LOT commits one to the view that representations of cows don't tell you anything about cows. On the contrary: having a mental symbol that refers to cows <u>presupposes</u> considerable knowledge about cows.

Perhaps it isn't so bad that LOT entails that the representations that are satisfied by cows have only an indirect role in the explanation of cow cognition, for there are always mental sentences to tell us about cows. But let's just be clear about what LOT is committed to here: The view we have arrived at is that cognition is essentially the application of a linguistically expressed theory. All the serious work gets done by sets of sentences that are internal tacit theories about whatever objects of cognition there happen to be. As far as cognizing cows goes, your |cow|s really don't matter; it is your tacit theory (or theories) of cows that does the work.

Enough has been said to suggest how the language of thought hypothesis provides for assembling representational structures from symbols whose semantic content is grounded in their functions as (portable) indicator signals. The strategy, however, has its limitations. Two are important for the present discussion. First, portable indicator signals that are assembled into complex representations, while a reality in digital computers, are surely not in the cards in the brain. As we have seen, complex LOT representations cannot have actual source-dependent indicator signals as constituents, for this would imply that every representation indicated the presence of the targets of each of its constituents. Such “representations”, indeed, would not be representations at all, but simply bundles of simultaneous feature detections. The transition from indication to representation in LOT systems is mediated by source-independent signal types that, severed from their source, denote the properties whose presence they detect when functioning as indicator signals.[11] But such source-<u>independent</u> (portable) indicator signals are neurally implausible, to say the least. What formal characteristics could distinguish the signal types? There is no evidence to support the idea that a distinctive spike train produced by a neural indicator retains its semantic significance when produced elsewhere in the brain. It is more plausible to suppose that a distinctive pattern of activation in a pool of associated neurons might retain its significance if copied, or simply repeated, elsewhere. But there is no evidence whatever for the suggestion that such patterns are not only the distinctive outputs of detection circuits, but are also assembled into complex representations. The fear, voiced early and often by connectionist critics of LOT systems, that LOT systems have no plausible neural implementation, seems well-founded.

The second limitation of LOT systems has to do with their restricted representational power. Representations in the language of thought have whatever structure is implicated by their combinatorial syntax. In language like schemes, the structure in question is logical form. While propositions arguably have structures isomorphic to logical forms, it is surely the case that many, perhaps most, representational targets of significance to cognitive systems have structures of entirely different sorts. Natural images of the sort studied by Olshausen and Field (1996), for example, certainly do not have logical forms, nor do problem spaces of the sort studied by planning theory in artificial intelligence. (Newell and Simon, 1972; See Cummins, 1996, for more on this theme.) Representational systems whose non-syncategorematic elements get their meanings by inheriting denotations from their roles in indication are, inevitably, denotational schemes, schemes whose semantics is the familiar truth-conditional semantics. The only thing such schemes can represent are propositions or complex propositional functions. Everything else is simply denoted. You can call denotation representation, provided you keep in mind the very real difference between something that merely labels its target, and something that actually provides information about it. Grounding representation in indication by promoting portable indicator signals into the semantic constituents of complex representations inevitably leads to a scheme that represents propositions and nothing else.

LOT schemes get around this limitation by encoding structure rather than representing it. To see the difference, compare LOT schemes for representing sentences in a natural or artificial language with Gödel numbering. A LOT scheme represents a target sentence S by tokening a sentence in the language of thought that has the same relevant structure — e.g., logical form — as S. In the Gödel numbering scheme, on the other hand, words are assigned natural numbers, and their positions in a sentence are encoded by prime numbers in ascending order. A word in position n assigned to m yields the number n^m. This number is uniquely factorable into n, m times. The Gödel number of a sentence is determined by multiplying all these uniquely factorable numbers together, yielding a uniquely factorable number from which the sentence can be recovered. For example, Assume that 'John', 'Mary' and 'loves' are assigned 2, 3 and 4 respectively. Then 'John loves Mary' is encoded as $2^2 \times 4^3 \times 3^5$ = 62208. This number may be represented in any convenient number system. The result is a numeral that does not itself share structure with the sentence it encodes. Information about the constituent structure of the sentence is still there, though well disguised, and this is what makes it possible to devise arithmetical processes that manipulate the relevant structural information without actually recovering it and representing it explicitly. An active-passive transformation, for example, can be written simply as an arithmetic procedure.[12]

4. The assembly of complex visual representations.

According to indicator-based theories of mental content generally, and to LOT theories in particular, mental representations either are indicator signals, or inherit their content from their roles, or the roles of their constituents, as indicator signals. We have seen however, that there are serious reasons for doubting that complex representations in the brain could be semantically composed of constituents whose meanings are inherited from their roles as indicator signals.

An entirely different relation between indication and representation emerges if we examine the way in which the sort of indicators discovered by Hubel and Weisel are implicated in the assembly of visual images. One account of this is to be found in recent research by Field and Olshausen (Field, 1987, 1994; Olshausen and Field, 1996, 1997, 2000).

Natural images contain much statistical structure and redundancies (Field 1987) and early visual processing functions to retain the information present in the visual signal while reducing the redundancies. In the 1950s, Stephen Kuffler (1952) discovered the center-surround structure of retinal ganglion cells' response and, 40 years later, Joseph Atick (1992) showed that this arrangement serves to decorrelate these cells' responses. As Horace Barlow (1961) had suspected, sensory neurons are organized to maximize the statistical independence of their response. Bruno Olshausen and David Field recently showed that the same is true of neurons in the primary visual cortex. Hubel and Wiesel discovered that neurons in the primary visual cortex are sensitive to edges, hence their functional description as edge detectors, but could only guess at the functional relevance of this structure (Hubel, 1988). According to Olshausen and Field, edge detection allows neurons in the primary visual cortex to respond to visual signals in a maximally independent fashion and thus produce sparsely coded representations of the visual field. To show this, they constructed an algorithm that could identify the minimal set of maximally independent basis functions capable of describing natural images (or small 12 by 12 pixel patches thereof) in a way that preserves all the information present in the visual signal. Because natural images tend to contain edges, and because there are reliable higher-order correlations (three-point and higher) between pixels along an edge, it turns out that natural images can be fully described with minimal resources as composites of about a hundred such basis functions (see figure 1). Given the statistical structure of natural images in the environment, there had to be such a set of functions, but the important point is that these basis functions are similar to those Hubel and Wiesel found 40 years earlier: spatially localized and oriented edges. Recently, O'Reilly and Munakata (2000) showed how to train a neural network using a form of Hebbian learning (conditional PCA) to produce a similar set of basis functions.

INSERT FIGURE 2 ABOUT HERE

To understand how visual representations are constructed out of these basis functions, consider a vector of V1 cortical cells[13] connected to the same retinal area via the same small subset of LGN cells[14]. Each cell has a receptive field similar to one in the minimal set of basis functions in Figure 1 above. Visual representations (of a small retinal area) can thus be thought of as a vector of cell activation, and an observed state of affairs (the inferred distal stimuli) as a linear function of a visual representation and some noise factor. A natural, but ultimately wrong, way to understand the construction of visual representations is as the activation of a subset of such basis functions *solely on the basis of the information each cell receives from the relevant retinal region*. In such a case, cells in the primary visual cortex would function as indicators of activity in the LGN and, ultimately, of properties in the visual field. Indeed, it would be easy to determine what state of affairs is currently observed if the basis functions were completely independent and the noise factor was known, but neither is the case. Because of this, many distinct visual representations (sets of basis vector functions) will account for the same observed state of affairs. Lewicki and Olshausen (1999) have shown, however, that it is possible to construct (infer) a unique visual representation given prior knowledge of the observed environment, that is, prior probabilities (the probability of observed states of affairs) and likelihood functions (the probability of visual representations given observed states of affairs). Instead of constructing a visual representation from a set of indicator signals, the visual system may infer the representation from indicator signals and relevant probabilistic knowledge of the visual environment.

Visual representations are thus constructed from retinal indicator signals, knowledge of the high-order correlational structure in the environment (coded in the LGN-V1 connections) and knowledge of relevant prior probabilities and likelihood functions. The picture that emerges here involves the construction of an image from a set of indicator signals that have the following characteristics:

- They are surprisingly few in number;

- They indicate multi-point correlations between adjacent pixels in the (whitened version of) the input;
- Their prowess as detectors of their proprietary targets is due, to a large extent, to recurrent circuitry that, in effect, computes a Bayesian function in which the prior probabilities are determined by the properties of neighboring areas of the developing image;
- Their representational content is semantically disjoint from that of the image they compose in the same way that pixel values themselves are semantically disjoint from the representational content of a computer graphic.

Like maps and scale models, image representations thus constructed have their meanings in virtue of their geometry rather than their origins. This is what gives them the source-independence characteristic of representations rather than indicator signals, and what allows for the possibility of disciplined structure sensitive transformations. Because such representations literally share structure with their targets, both static and dynamic structural properties of those targets can be mirrored by learning to transform the representations in ways that mirror the ways in which nature constrains the structure of, and structural changes in, the targets. Faces can be aged, objects can be rotated or "zoomed", projections into three dimensions can be computed. None of this is thinkable in a system in which visual representations are <u>semantically</u> composed from constituents whose contents are determined by their roles as indicator signals.

5. Representations and Targets

We have been emphasizing the source-independence that representations achieve in virtue of their distinctive internal structure which either mirrors or encodes the structure of what is represented, or encodes semantic combinatorics. To be useful, however, representations must typically retain a kind of source-dependence. To understand this, we need to understand the relation between a representation and its target.

“x represents y” belongs to a group of relational semantic terms that are ambiguous. This can be seen by reflecting on a comment of Jerry Fodor’s. “Misrepresentation,” Fodor said, “is when a representation is applied to something it doesn’t apply to.” (Email exchange.) It is obvious that ‘applies to’ has to be ambiguous if this remark is to make sense and be true, which it evidently does and is. Analogous observations apply to “x refers to y”, “x is true of y”, “x means y” and “x represents y”. For example, you say, "I used that map and found my way around the city with no problem." "Which city do you mean?" ("Which city are you referring to?") I ask. Here, I am asking for your <u>target</u>, the city against which that map’s accuracy is to be measured. "Here is the city map I was telling you about," you say. "Which city do you mean?" ("Which city are you referring to?") I ask. Here, I am asking for the <u>content</u>, i.e., for the city the map actually represents—the city with which the map shares structure.

Though unmarked in ordinary vocabulary, the distinction between representational targets and contents is a commonplace. We are looking through my pictures of opera singers. You ask me what Anna Moffo looked like in her prime, and I hand you a picture, which happens to be a picture of Maria Callas. (It’s hard to believe, but it is just a Philosophy example.) We can distinguish here between the target of my representing —

Anna Moffo, the singer the picture I handed you was supposed to represent on this occasion — and the content of the picture I produced — Maria Callas. The target of a representation on a particular occasion of its use is whatever it is supposed to represent on that occasion of use. The content of a representation is whatever it actually does represent.

So: a representation can be applied to something it doesn't apply to because there can be a mismatch between the content of a representation and its target on a particular occasion of its use, a mismatch between what it actually represents — its representational content — and what it is intended to represent on that occasion. Let's call something whose function is to represent some target or class of targets an <u>intender</u>. Indenters come in many forms. A clear example would be the process hypothesized by standard speech-recognition models whose function is representing the phrase structure of the current linguistic input. This process has a proprietary class of targets, <u>viz</u>., phrase structures of the current linguistic inputs. The accuracy of the representational performances of that process are to be judged by reference to the degree of match between the current target — the actual phrase structure of the current input — and the actual (as opposed to intended) content of the representation it produces on that occasion.

A representation's target, as opposed to its representational content, is source-dependent. Processes downstream of the parser lately imagined — its clients — operate on the assumption that the representations it gets from that intender are representations of the phrase structure of the current linguistic input. To know that a token representation's target is the phrase structure of the current linguistic input amounts to knowing it was produced by an intender whose business it is to represent the phrase structure of the current linguistic input. If the representation were produced by a different intender, one whose business it is to represent move-trees in checkers, for example, it would need very different treatment (though it probably wouldn't get it in this case).

Exactly this sort of source-dependency is exhibited by the visual representations built up in V1. Consumers of these representations other than mere enhancers need (typically) to operate on the assumption that they are processing representations of the current visual scene, and not, say, a remembered scene. This information is not encoded in the structure of the representation itself. To be available to a consumer that has more than one source, then, the information has to be carried by the fact that a particular intender produced the representation in question.

It is tempting to suppose that source information, hence target information, could be incorporated into the representation itself. To see the limitations of this idea, imagine labeling year book photographs with the names of the people represented. This gives us representations with contents of the form, "Susan Smith looks like this <photo>." There is nothing wrong with this scheme, provided consumers know that the targets of the complex representations — name plus picture — are true propositions about people's appearances.[15] But imagine someone using such a representation in a reductio argument. "Suppose Maria Callas looked like this <photo>," she starts out, exhibiting a photo of Anna Moffo. Her target is a false proposition about how Maria Callas looked. A consumer who does not realize this will get things seriously wrong. Labeling a photo just produces a new complex representation with a new content. Its target could be anything. Though most things, e.g., the number nine, will be non-starters as the target, it doesn't follow that they couldn't be the target. It is just unlikely that any intender would be that stupid. A representation cannot say what its own target is.[16]

Thus, representations have this in common with indicator signals: just as the information carried by an indicator signal depends on which indicator produces the signal, so the information carried by a token representation depends, in part, on the intender that produces it. Unlike indicator signals, however, representations carry lots of information about their targets beyond the fact that they are present. And this is precisely what makes them useful.[17]

6. Accuracy

There are two different ways in which indication can be accurate or inaccurate. Simple binary indicators of the "idiot light" variety are either on or off, hence either right or wrong. An indicator signal in such a case cannot be more or less accurate. But the indicator that produces it may be said, somewhat misleadingly, to be more or less accurate as a function of its reliability.

Indicator signals often come in degrees, however. Thermometers can fail to be accurate because they have the amount of something as a target, and they can get the amounts wrong by a greater or lesser margin. The cells studied by Hubel and Weisel present an interesting case of this kind. A cell firing at less than peak but more than is involved in normal background noise, could be interpreted as indicating (1) a probability that its target is present, or (2) how close the current stimulus is to the target. Both interpretations allow for individual signals having a degree of accuracy that is independent of the cell's reliability, since both involve indicating the amount of something (probability, nearness). On the other hand, it may be that less than peak firing is simply ignored, in which case the cells in question are, essentially, binary indicators.

Indicators say only whether (or how much of) their target is there; they say nothing about what their targets are like. Representations do carry information about what their targets are like, and this makes representational accuracy a far more complex affair than indicator accuracy.

Representational accuracy is a relation between representation and a target. Once we recognize that accuracy is a relation between a representation and its target on a particular occasion of its deployment, it becomes clear that representations are not accurate or inaccurate in their own right. What we are calling accuracy thus differs sharply from truth. A sentence may express a true proposition, yet be an inaccurate representation of its target. This is precisely the situation in reductio arguments, where the proposition targeted by the sentence expressing the supposition is (if the argument is successful) a false proposition. Similarly, if you are asked to specify some false proposition about the Eiffel Tower, your performance will be accurate only if you express a false proposition. In general, targets need not be actual states of affairs, objects, events, or whatever. The target a particular token representation is aimed at is fixed by the function of the intender that tokens that representation, together with the facts that happen to obtain at the time. Thus, an intender may have the function of representing the phrase structure of the current linguistic input, and that function, together with the actual (phrase) structure of the then current input will determine what the target of a particular token happens to be, and hence the standard against which the accuracy of the representation is to be measured.

Propositions are peculiar targets in that they cannot be represented with greater or less accuracy: they are either hit, or missed. There is no such thing as varying degrees of accuracy when it comes to representing the proposition that the Eiffel Tower is in Paris. Someone might say it is in France, and claim that that is “closer” than saying that it is in Europe or Australia. But this is evidently not a matter of getting closer to expressing the right proposition, but a matter of specifying a location that is closer to the correct one. Proposition expressing, at least as the going theory has things, is an all-or-nothing affair (Fodor and LePore, 1992). But most representation isn’t like this. Pictures, maps, diagrams, graphs, scale models, and, or course, partitioned activations spaces, are more or less accurate representations of the targets they are applied to.

It follows from this observation that most representations cannot have propositional contents. It might seem that pictures could express propositions, because they could be said to hold in some possible worlds and not in others. It seems that one could take a video tape of a convenience store robbery, and alter it in various ways so that the variants held in possible worlds that were “close” to the actual world, worlds in which everything went down just as it did in the actual world except that the perpetrator had a moustache, or the clock on the wall behind the clerk said ten twenty-five instead of ten twenty-two. Since a proposition can be conceived as a set of possible worlds (Stalnaker 1984), it might seem that a picture could be regarded as expressing the proposition that consists in the set of possible worlds it actually depicts accurately.

But no picture depicts with perfect accuracy. This is not just a consequence of rendering three dimensions in two. One might reasonably hold that the target of a photograph, for example, is not the three dimensional spatial layout, but its two-dimensional projection at a certain plane. Even granting this, however, there is still the fact that getting the color right often requires inaccurate illumination and a sacrifice in resolution. Depth of field issues will inevitably render some things sharper than others. A video that gets the perpetrator's face in focus but blurs the clock on the wall behind him misrepresents the scene as having a blurry clock. We are not fooled, of course: we know clocks in convenience stores are not blurry. This makes it tempting to suppose that the photo doesn't really represent the clock as blurry. But it does: a photo of things we don't antecedently know about — one taken through a microscope, for example — can leave us wondering whether we have an accurate picture of a blurry object or a depth of field problem. Similarly, there are many compromises involved in shooting moving subjects with still shots, and even with movies or video. When we take all this into account, we no longer have a set of possible worlds corresponding to a given picture, but a smear of more or less similar worlds spreading out along various dimensions. Is the photo that nails the color of Aunt Tilly's hat while, inevitably, overestimating the intensity of illumination and blurring the features of the man running past her, more or less accurate than the one that allows us to identify the man but not the hat?[18]

Trade-offs of the sort illustrated by the depth of field problem in photographs are ubiquitous in non-symbolic representational schemes. Such schemes often represent many quantitatively variable properties simultaneously. Photos and models, for example, simultaneously represent color, relative distances, and sizes. It frequently happens that increasing accuracy in one dimension entails sacrificing accuracy in another. In human vision, sensitivity is in conflict with resolution and color accuracy because of the high density of the relatively insensitive cones at the fovea. (This is also why color blind individuals have better night vision.) Sentences (again, as construed by the tradition of truth-conditional semantics) can get around problems like this by abstracting away from some properties while focusing on others, and by abstracting away from troublesome differences in degree. “Roma tomatoes are red when ripe,” is simply silent on size, shape, and variation in shade, intensity and saturation. A photo or model cannot do this, nor can biological vision systems.

For those of us brought up to think of semantics in a linguistic setting, the striking thing about maps, diagrams, partitioned activations spaces, images, graphs and other non-linguistic representations is that they are not true or false, and that their accuracy comes in degrees. A sentence either hits its propositional target, or it fails. Non-propositional representations require a graded notion of accuracy. Moreover, such representations are typically multi-dimensional. Images, for example, represent (relative) size, shape, color and (relative) location simultaneously. The possibility thus arises that two image representations might be incomparable in over-all accuracy, since one might do better on some dimensions — size and shape, say — while the other does better on others — color and location.[19] The fact that non-propositional representations can simultaneously represent along many dimensions probably precludes any sort of “all things considered” or “over all” accuracy scores. The concepts of truth and falsehood, and the Tarskian combinatorial semantics we have come to associate with them, will be no help at all in understanding how these non-propositional representations fit or fail to fit their targets. Representational meaning for non-propositional representations will have to be understood in different terms, as will their semantic structures. It is precisely this rich area of inquiry that is opened up when we distinguish representation from indication, and contemplate mechanisms other than those presupposed by the Language of Thought for assembling representations from indicator signals.

NOTES

[1] This way of putting things—that believing that it will rain and desiring that it will rain differ only in that a mental representation meaning that it is raining is in the Belief Box in the first case and in the Desire Box in the second—is due to Schiffer (1987).

[2] Conceptual role theories of mental content have difficulty distinguishing mental representations from the attitudes because they take the mental content of x to be a function of x' s epistemic liaisons. Propositional attitudes have epistemic liaisons, but representations, as we use the term, do not. See Cummins, 1996, chapter 3 for a detailed discussion of this point.

[3] The cells were discovered by David Hubel and Torsten Wiesel (1962). They describe the behaviors of these cells this way: "The most effective stimulus configurations, dictated by the spatial arrangements of excitatory and inhibitory regions, were long narrow rectangles of light (slits), straight-line borders between areas different brightness (edges), and dark rectangular bars against a light background."

[4] Thermostats using bimetallic strips are usually designed so that when the bimetallic changes shape “enough”, it closes a circuit that turns on the furnace, or air conditioner, or whatever. The electrical pulse that closes the furnace relay is also an indicator signal, but it doesn’t indicate the temperature, except relative to a fixed thermostat setting.

[5] The theory is generally credited to Dennis Stampe (1977). Its most prominent advocates are Fodor (1987, 1990) and Dretske (1981, 1988).

[6] It is also what Haugeland would call a recording of the picture. See Haugeland (1991).

[7] Actually, a structural encoding (Cummins *et al*., 2001) of the information a representation would carry would do the trick as well. This does not affect the present point, which is that indicator signals would <u>not</u> do the trick.

[8] The structural properties of an indicator signal might be used to type-identify it. More of this shortly.

[9] We do not mean to imply here that the shape of a spike train is never significant. The point is rather that two indicators can have the same spike train, yet indicate different things.

[10] Qualifications are required to turn this into an even mildly plausible definition of content. See Cummins (1989) for details.

[11] Or rather: they are, when things are working properly, processed as if they denoted the properties they detect when functioning accurately as indicator signals.

[12] For an extended discussion of this theme, see Cummins *et al*., 2001.

[13] In reality, this vector is contained within a hypercolumn, which also contains a similar contralateral vector (for the other eye). All cells in a given hypercolumn respond to the same region of retinal space. Neighboring hypercolumns respond to neighboring regions of retinal space (Livingstone and Hubel, 1988).

[14] The LGN (the lateral geniculate nucleus of the thalamus) is not a mere relay station en route to the cortex since it performs important computations of its own. But these will not matter here, as they are mostly concerned with the dynamic, as opposed to structural, aspects vision and, owing to the massive backprojections from the cortex to LGN, with visual attention processes (it is estimated that there are ten times more packprojections from the cortex to the LGN as there are projections from the LGN to the cortex, Sherman and Koch 1986).

[15] It actually isn't clear that the targets could be propositions, since the representations have pictures as constituents. See Cummins, 1999, for an argument against the view that pictures could have propositional contents. From the fact that pictures cannot have propositional contents, it doesn't <u>follow</u> that representations with pictures as constituents cannot have propositional contents. But it makes it seem problematic.

[16] This is why LOTers put LOT sentences into boxes like the Belief Box. An **|I am rich|** in the Belief Box is presumed to have a true proposition as target. Not so an **|I am rich|** in the Desire Box.

[17] There is much more to be said about intenders and target fixation. Some of this can be found in Cummins (1996 and 2000 Millikan reply).

[18] We could, perhaps, imagine a possible world whose natural laws dove-tailed with the constraints of a given representational scheme, so that, for example, things flattened into two dimensions when photographed, with objects originally at different distances losing various degrees of resolution themselves. These would be radically counter-nomic worlds. The laws of our world, and every world satisfying anything like the same laws, preclude perfect accuracy for most representational schemes.

[19] It seems likely that high accuracy on one dimension will often have to be paid for in lower accuracy in others, given limited resources. The eye, for example, gains considerable resolution and color information via foveation, but loses light sensitivity in the process. A map that shows all the streets of London on one page will be either too big to use in the car, or be viewable only with magnification.

REFERENCES

Atick, J. (1992). "Could information theory provide an ecological theory of sensory processes?" Network 3:213-251.

Barlow, H.B. (1961). "Possible principles underlying the transformation of sensory messages." W.A. Rosenblueth (ed.) (1961). *Sensory Communication*. Cambridge, Mass.: MIT Press, pp. 217-234.

Churchland, P.M. (1995). *The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain* Cambridge, Mass.: MIT Press.

Cummins, R. (1989). *Meaning and Mental Representation*. Cambridge, Mass.: MIT Press.

Cummins, R. (1996). *Representations, Targets, and Attitudes*. Cambridge, Mass.: MIT Press.

Cummins, R. (1999). “Truth and Meaning.” J. Keim Campbell, M. O’Rourke and D. Shier (eds.) *Meaning and Truth: Investigations in Philosophical Semantics*. New York: Seven bridges Press.

Cummins, R., J. Blackmon, D. Byrd, P. Poirier, M. Roth, G. Schwarz (2001). “Systematicity and the Cognition of Structured Domains.” *The Journal of Philosophy* 98: 1-19.

Dretske, F.I. (1981). *Knowledge and the Flow of Information*. Cambridge, Mass.: MIT Press.

Dretske, F.I. (1988). *Explaining Behavior: Reasons in a World of Causes*. Cambridge, Mass.: MIT Press.

Field, D.J. (1987). “Relations between the statistics of natural images and the response properties of cortical cells.” *Journal of the Optical Society of America, A*, 4:2379-2394.

Field, D.J. (1994). “What is the goal of sensory coding?” *Neural Computation* 6:559-601.

Fodor, J.A. (1975). *The Language of Thought*. New York: T.Y. Crowell.

Fodor, J.A. (1987). *Psychosemantics*. Cambridge, Mass.: MIT Press.

Fodor, J.A. (1990). *A Theory of Content and other essays*. Cambridge, Mass.: MIT Press.

Fodor, J.A. and E. LePore (1992). *Holism: A Shopper's Guide*. Oxford: Blackwell.

Haugeland, J. (1991). Representational genera. In W. Ramsey, S. Stich, and D. Rumelhart (eds.) *Philosophy and Connectionist Theory*. Hillsdale (NJ): Lawrence Erlbaum.

Hubel, D. H. and T. N. Wiesel (1962). “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex”. *Journal of Physiology*. 160:106-154.

Hubel, D.H. (1988). *Eye, Brain, and Vision*. New York: Scientific American Library.

Kuffler, S. (1952). “Neurons in the retina: Organization, Inhibition and excitatory problems.” *Cold Spring Harbor Symposia on Quantitative Biology*. 17: 281-292.

Lewicki, M.S. and B.A. Olshausen (1999). “A Probabilistic framework for the adaptation and comparison of images codes.” *Journal of the Optical Society of America, A*, 16:1587-1601.

Livingstone, M. and D. Hubel (1988). “Segregation of from, color, movement and depth: Anatomy, physiology, and perception.” *Science* 240:740-749.

Millikan, R.G. (2000) “Representations, Targets, Attitudes.” *Philosophy and Phenomenological Research*. 60, 103-111.

Newell, A. and H. Simon (1972). *Human Problem Solving*. Englewood Cliffs, N.J., Prentice-Hall

Olshausen, B.A. and D.J. Field (1996). “Emergence of simple-cell receptive field properties by learning a sparse code for natural images.” *Nature* 381:607-609.

Olshausen, B.A. and D.J. Field (1997). “Sparse coding with an overcomplete basis set: A strategy employed by V1?” *Vision Research* 37:3311-3325.

Olshausen, B.A. and D.J. Field (2000). “Vision and the coding of natural images.” *American Scientist* 88:238-245.

O’Reilly R.C. and Y. Munakata (2000). *Computational Explorations in Cognitive Neuroscience*. Cambridge, Mass.: MIT Press.

Schiffer, S. (1987). *Remnants of Meaning*. Cambridge, Mass.: MIT Press.

Sherman, S.M. and C. Koch (1986). “The control of retino-geniculate transmission in the mammalian lateral geniculate nucleus.” *Experimental Brain Research* 63:1-20.

Stalnaker, R. (1984). *Inquiry*. Cambridge, Mass.: MIT Press.

Stampe, D. (1977). “Towards a causal theory of linguistic representation.” *Midwest Studies in Philosophy* 2:42-63.

FIGURE 1

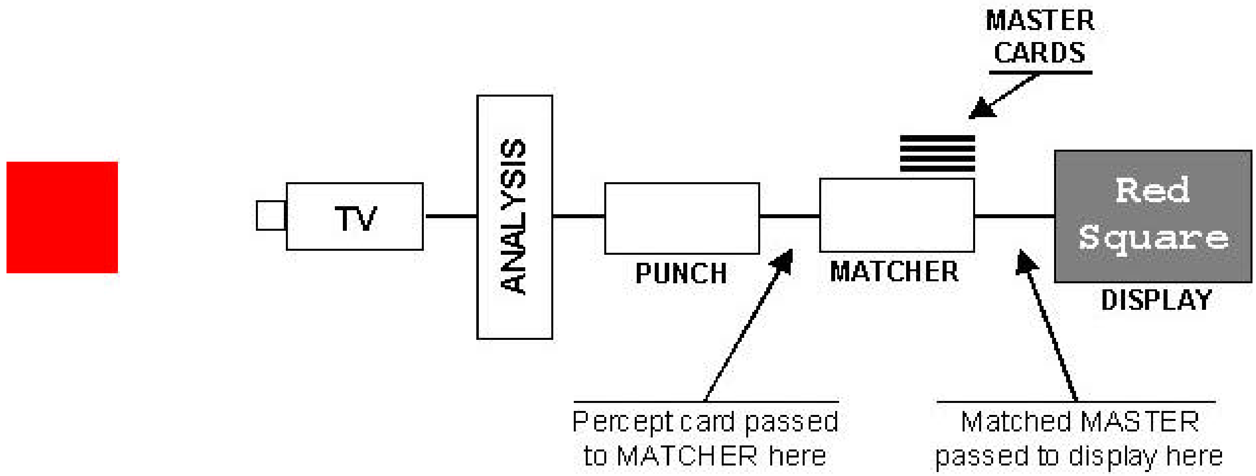

Figure 1. The Locke machine (from Cummins 1989).

FIGURE 2

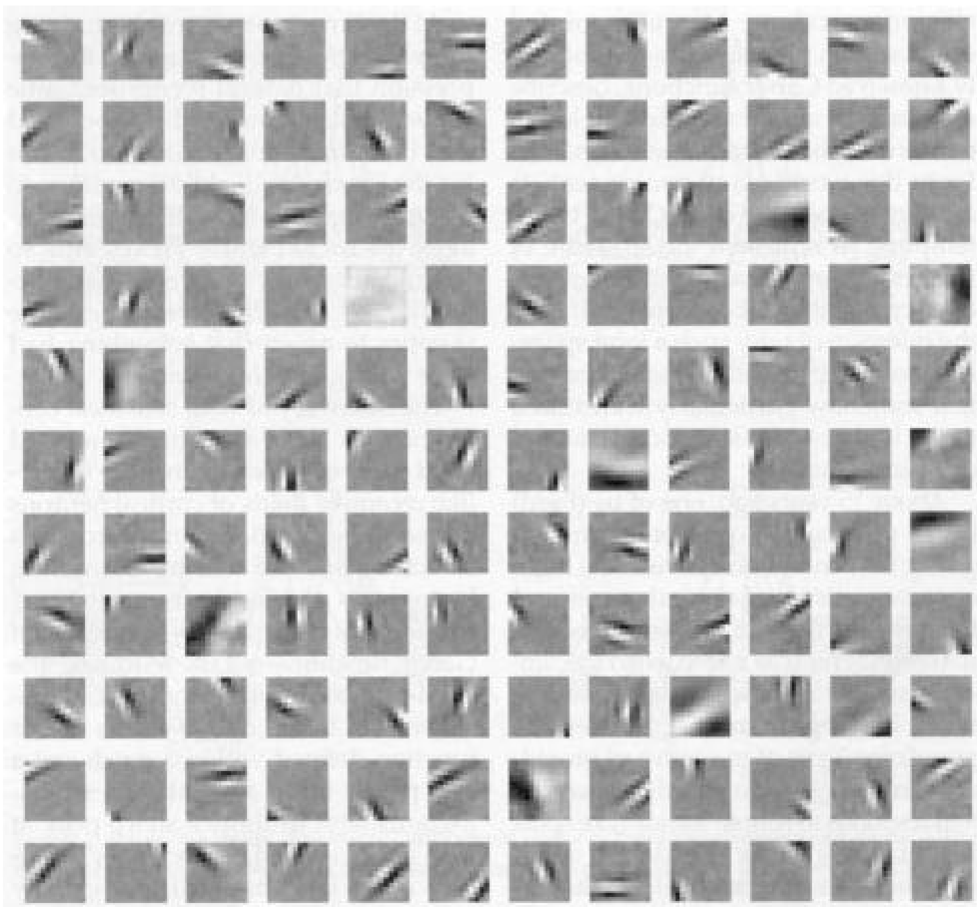

Figure 2. Optimal basis function set to represent natural images (from Olshausen and Field 2000).