

Paper presented at UQàM Summer Institute in Cognitive Sciences on *Categorisation* 2003  
<http://www.unites.uqam.ca/sccog/liens/program.html>  
To appear in Lefebvre C., & H. Cohen (Eds.) (2005) *Handbook on Categorization*. Elsevier

## To Cognize is to Categorize: Cognition is Categorization

Stevan Harnad  
Chaire de recherche du Canada  
Centre de neuroscience de la cognition  
Université du Québec à Montréal

**ABSTRACT:** We organisms are sensorimotor systems. The things in the world come in contact with our sensory surfaces, and we interact with them based on what that sensorimotor contact “affords”. All of our categories consist in ways we behave differently toward different *kinds* of things -- things we do or don’t eat, mate-with, or flee-from, or the things that we describe, through our language, as prime numbers, affordances, absolute discriminables, or truths. That is all that cognition is for, and about.

**KEYWORDS:** abstraction, affordances, categorical perception, categorization, cognition, discrimination, explicit learning, grounding, implicit learning, invariants, language, reinforcement learning, sensorimotor systems, supervised learning, unsupervised learning

*Pensar es olvidar diferencias, es generalizar, abstraer.*  
En el abarrotado mundo de Funes no había sino detalles, casi inmediatos. Borges (“Funes el memorioso”)

**1. Sensorimotor Systems.** Organisms are sensorimotor systems. The things in the world come in contact with our sensory surfaces, and we interact with them based on what that sensorimotor contact “affords” (Gibson 1979).

**2. Invariant Sensorimotor Features (“Affordances”).** To say this is not to declare oneself a “Gibsonian” (whatever that means). It is merely to point out that what a sensorimotor system can do is determined by what can be extracted from its motor interactions with its sensory input. If you lack sonar sensors, then your sensorimotor system cannot do what a bat’s can do, at least not without the help of instruments. Light stimulation affords color vision for those of us with the right sensory apparatus, but not for those of us who are color-blind. The geometric

fact that, when we move, the “shadows” cast on our retina by nearby objects move faster than the shadows of further objects means that, for those of us with normal vision, our visual input affords depth perception.

From more complicated facts of projective and solid geometry it follows that a 3-dimensional shape, such as, say, a boomerang, can be recognized as being the same shape – and the same size – even though the size and shape of its shadow on our retinas changes as we move in relation to it or it moves in relation to us. Its shape is said to be invariant under these sensorimotor transformations, and our visual systems can detect and extract that invariance, and translate it into a visual constancy. So we keep seeing a boomerang of the same shape and size even though the shape and size of its retinal shadows keep changing.

**3. Categorization.** So far, the affordances I’ve mentioned have depended on having either the right sensors, as in the case of sonar and color, or the right invariance-detectors, as in the case of depth perception and shape/size constancy. Having the ability to detect the stimulation or to detect the invariants in the stimulation is not trivial; this is confirmed by the fact that sensorimotor robotics and sensorimotor physiology have so far managed to duplicate and explain only a small portion of this subset of our sensorimotor capacity. But we are already squarely in the territory of categorization here, for, to put it most simply and generally: categorization is any systematic differential interaction between an autonomous, adaptive sensorimotor system and its world: Systematic, because we don’t want arbitrary interactions like the effects of the wind blowing on the sand in the desert to be counted as categorization (though perhaps there are still some inherent similarities there worth noting). Neither the wind nor the sand is an autonomous sensorimotor system; they are, jointly, simply dynamical systems, systems that interact and change according to the laws of physics.

Everything in nature is a dynamical system, of course, but some things are not only dynamical systems, and categorization refers to a special kind of dynamical system. Sand also interacts “differentially” with wind: Blow it this way and it goes this way; blow it that way and it goes that way. But that is neither the right kind of systematicity nor the right kind of differentiability. It also isn’t the right kind of adaptivity (though again, categorization theory probably has a lot to learn from ordinary dynamical interactions too, even though they do not count as categorization).

Dynamical systems are systems that change in time. So it is already clear that categorization too will have to have something to do with changes across time. But adaptive changes in autonomous systems are those in which internal states within the autonomous system systematically

change with time, so that, to put it simply, the exact same input will not produce the exact same output across time, every time, the way it does in the interaction between wind and sand (whenever the wind blows in exactly the same direction and the sand is in exactly the same configuration). Categorization is accordingly not about exactly the same output occurring whenever there is exactly the same input. Categories are kinds, and categorization occurs when the same output occurs with the same kind of input, rather than the exact same input. And a different output occurs with a different kind of input. So that's where the "differential" comes from.

**4. Learning.** The adaptiveness comes in with the real-time history. Autonomous, adaptive sensorimotor systems categorize when they respond differentially to different kinds of input, but the way to show that they are indeed adaptive systems -- rather than just akin to very peculiar and complex configurations of sand that merely respond (and have always responded) differentially to different kinds of input in the way ordinary sand responds (and has always responded) to wind from different directions -- is to show that at one time it was not so: that it did not always respond differentially as it does now. In other words (although it is easy to see it as exactly the opposite): categorization is intimately tied to learning.

Why might we have seen it as the opposite? Because if instead of being designers and explainers of sensorimotor systems and their capacities we had simply been concerned with what kinds of things there are in the world, we might have mistaken the categorization problem as merely being the problem of identifying what it is that exists (that sensorimotor systems can then go on to categorize). But that is the ontic side of categories, concerned with what does and does not exist, and that's probably best left to the respective specialists in the various kinds of things there are (specialists in animals, vegetables, or minerals, to put it simply). The kinds of things there in the world are, if you like, the sum total of the world's potential affordances to sensorimotor systems like ourselves. But the categorization problem is not determining what kinds of things there are, but how it is that sensorimotor systems like ourselves manage to detect those kinds that they can and do detect: how they manage to respond differentially to them.

**5. Innate Categories.** Now it might have turned out that we were all born with the capacity to respond differentially to all the kinds of things that we do respond to differentially, without ever having to learn to do so (and there are some, like Jerry Fodor (1975, 1981, 1998), who sometimes write as if they believe this is actually the case). Learning might all be trivial; perhaps all the invariances we can detect, we could already detect

innately, without the need of any internal changes that depend on time or any more complicated differential interaction of the sort we call learning.

This kind of extreme nativism about categories is usually not far away from something even more extreme than nativism, which is the view that our categories were not even “learned” through evolutionary adaptation: The capacity to categorize comes somehow prestructured in our brains in the same way that the structure of the carbon atom came prestructured from the Big Bang, without needing anything like “learning” to shape it. (Fodor’s might well be dubbed a “Big Bang” theory of the origin of our categorization capacity.)

Chomsky (e.g., 1976) has made a similar conjecture – about a very special subset of our categorization capacity, namely, the capacity to generate and detect all and only those strings of words that are grammatical according to the Universal Grammar (UG) underlying all possible natural languages: UG-compliance is the underlying invariant in question, and, according to Chomsky, our capacity to detect and generate UG-compliant strings of words is shaped neither by learning nor by evolution; it is instead somehow inherent in the structure of our brains as a matter of structural inevitability, directly from the Big Bang. This very specific theory, about UG in particular, is not to be confused with Fodor’s far more general theory that all categories are unlearned and unevolved; in the case of UG there is considerable “poverty-of-the-stimulus” evidence to suggest that UG is not learnable by children on the basis of the data they hear and produce within the time they take to learn their first language; in the case of most of the rest of our categories, however, there is no such evidence.

**6. *Learned Categories.*** All evidence suggests that most of our categories are learned. To get a sense of this, open a dictionary at random and pick out a half dozen “content” words (skipping function words such as “if,” “not” or “the”). What you will find is nouns, verbs, adjectives and adverbs all designating categories (kinds of objects, events, states, features, actions). The question to ask yourself is: Was I born knowing what are and are not in these categories, or did I have to learn it?

You can also ask the same question about proper names, even though they don’t appear in dictionaries: Proper names name individuals (e.g., people, places) rather than kinds, but for a sensorimotor system, an individual is effectively just as much of a kind as the thing a content word designates: Whether it is Jerry Fodor or a boomerang, my visual system still has to be able to sort out which of its shadows are shadows of Jerry Fodor and which are shadows of a boomerang. How?

**7. Supervised Learning.** Nor is it all as easy as that case. Consider the more famous and challenging problem of sorting newborn chicks into males and females. I'm not sure whether Fodor thinks this capacity could be innate, but the grandmaster, 8<sup>th</sup>-degree black-belt chicken-sexers on this planet – of which there are few, most of them in Japan – say that it takes years and years of trial and error training under the supervision of masters to reach black-belt level; there are no short-cuts, and most aspirants never get past brown-belt level. (We will return to this.) Categorization, it seems, is a sensorimotor skill, though most of the weight is on the sensory part (and the output is usually categorical, i.e., discrete, rather than continuous); and, like all skills, it must be learned.

So what is learning? It is easier to say what a system does when it learns than to say how it does it: Learning occurs when a system samples inputs and generates outputs in response to them on the basis of trial and error, its performance guided by corrective feedback. Things happen, we do something in response; if what we did was the right thing, there is one sort of consequence; if it was the wrong thing there is another sort of consequence. If our performance shows no improvement with time, then we are like the sand in the wind. If our performance improves – more correct outputs, fewer errors – then we are learning. (Note that this presupposes that there is such a thing as an error, or miscategorization: No such thing comes up in the case of the wind, blowing the sand.)

This sketch of learning should remind us of BF Skinner, behaviorism; and schedules of reward and punishment (Catania & Harnad 1988). For it was Skinner who pointed out that we learn on the basis of feedback from the consequences of our behavior. But what Skinner did not provide was the internal mechanism for this sensorimotor capacity that we and so many of our fellow-creatures have, just as Gibson did not provide the mechanism for picking up affordances. Both these thinkers thought that providing internal mechanisms was either not necessary or not the responsibility of their discipline. They were concerned only with describing the input and the sensorimotor interactions, not how a sensorimotor system could actually do those things. So whereas they were already beginning to scratch the surface of the “what” of our categorization capacity, in input/output terms, neither was interested in the “how.”

**8. Instrumental (Operant, Reinforcement) Learning.** Let us, too, set aside the “how” question for the moment, and note that so-called operant or instrumental learning -- in which, for example, a pigeon is trained to peck at one key whenever it sees a black circle and at another key whenever it sees a white circle (with food as the feedback for doing

the right thing and no-food as the feedback for doing the wrong thing) – is already a primitive case of categorization. It is a systematic differential response to different kinds of input, performed by an autonomous adaptive system that responded randomly at first, but learned to adapt its responses under the guidance of error-correcting feedback (thanks, presumably, to some sort of adaptive change in its internal state).

The case of black vs. white is relatively trivial, because the animal's sensory apparatus already has those two kinds of inputs well-segregated in advance -- although if, after training on just black and white, we began to "morph" them gradually into one another as shades of gray, and tested those intermediate shades without feedback, the pigeon would show a smooth "generalization gradient," pecking more on the "black" key the closer the input was to black, more on the white key the closer the input was to white, and approaching a level of chance performance midway between the two. The same would be true for a human being in this situation.

**9. Color Categories.** But if the animal had color vision, and we used blue and green as our inputs, the pattern would be different. There would still be maximal confusion at the blue-green midpoint, but on either side of that boundary the correct choice of key and the amount of pressing would increase much more abruptly – one might even say "categorically" -- than with shades of gray. The reason is that between black and white there is no innate category boundary, whereas between green and blue there is (in animals with normal green/blue color vision). The situation is rather similar to hot and cold, where there is a neutral point midway between the two poles, feeling neither cold nor hot, and then a relatively abrupt qualitative difference between the "warm" range and the "cool" range in either direction.

**10. Categorical Perception.** This relatively abrupt perceptual change at the boundary is called "categorical perception" (CP) and in the case of color perception, the effect is innate. Light waves vary in frequency. We are blind to frequencies above red (infrared, wavelength about 800 nm) or below violet (ultraviolet, wavelength about 400 nm), but if we did not have color CP then the continuum from red to violet would look very much like shades of gray, with none of those qualitative "bands" separated by neutral mixtures in between that we all see in the rainbow or the spectrum.

Our color categories are detected by a complicated sensory receptor mechanism, not yet fully understood, whose components include not just light frequency, but other properties of light, such as brightness and

saturation, and an internal mechanism of three specialized detectors selectively tuned to certain regions of the frequency spectrum (red, green, and blue), with a mutually inhibitory “opponent-process” relation between their activities (red being opposed to green and blue being opposed to yellow). The outcome of this innate invariance extracting mechanism is that some frequency ranges are automatically “compressed”: we see them all as just varying shades of the same qualitative color. These compressed ranges are then separated from adjacent qualitative regions, also compressed, by small, boundary regions that look like indefinite mixtures, neutral between the two adjacent categories. And just as there is compression within each color range, there is expansion between them: Equal-sized frequency differences look much smaller and are harder to detect when they are within one color category than when they cross the boundary from one category to the other (Berlin & Kay 1969; Harnad 2003).

Although basic color CP is inborn rather than a result of learning, it still meets our definition of categorization because the real-time trial-and-error process that “shaped” CP through error-corrective feedback from adaptive consequences was Darwinian evolution. Those of our ancestors who could make rapid, accurate distinctions based on color out-survived and out-reproduced those who could not. That natural selection served as the “error-correcting” feedback on the genetic trial-and-error variation. There are probably more lessons to be learned from the analogy between categories acquired through learning and through evolution as well as from the specific features of the mechanism underlying color CP -- but this brings us back to the “how” question raised earlier, to which we promised to return.

**11. Learning Algorithms.** Machine learning algorithms from artificial intelligence research, genetic algorithms from artificial life research and connectionist algorithms from neural network research have all been providing candidate mechanisms for performing the “how” of categorization.

There are in general two kinds of learning models: so-called “supervised” and “unsupervised” ones. The unsupervised models are generally designed on the assumption that the input “affordances” are already quite salient, so that the right categorization mechanism will be able to pick them up on the basis of the shape of the input from repeated exposure and internal analysis alone, with no need of any external error-correcting feedback.

By way of an exaggerated example, if the world of shapes consisted of nothing but boomerangs and Jerry-Fodor shapes, an unsupervised

learning mechanism could easily sort out their retinal shadows on the basis of their intrinsic structure alone (including their projective geometric invariants). But with the shadows of newborn chick abdomens, sorting them out as males and females would probably need the help of error-corrective feedback. Not only would the attempt to sort them on the basis of their intrinsic structural landscape alone be like looking for a needle in a haystack, but there is also the much more general problem that the very same things can often be categorized in many different ways. It would be impossible, without error-correcting supervision, to determine which way was correct in a given context. For the right categorization can vary with the context: sometimes we may want to sort baby chicks by gender, sometimes by species, sometimes by something else (Harnad 1987).

In general, a nontrivial categorization problem will be “underdetermined.” Even if there is only one correct solution, and even if it can be found by an unsupervised mechanism, it will first require a lot of repeated exposure and processing. The figure/ground distinction might be something like this: How, in general, does our visual system manage to process the retinal shadows of real-world scenes in such a way as to sort out what is figure and what is ground? In the case of ambiguous figures such as Escher drawings there may be more than one way to do this, but in general, there is a default way to do it that works, and our visual systems usually manage to find it quickly and reliably for most scenes. It is unlikely that they learned to do this on the basis of having had error-corrective feedback resulting from sensorimotor interactions with samples of the endless possible combinations of scenes and their shadows.

**12. Unsupervised Learning.** There are both morphological and geometric invariants in the sensory shadows of objects, highlighted especially when we move relative to them or vice versa; these can be extracted by unsupervised learning mechanisms that sample the structure and the correlations (including covariance and invariance under dynamic sensorimotor transformations). Such mechanisms cluster things according to their structural similarities and dissimilarities, enhancing both the similarities and the contrasts. An example of an unsupervised contrast-enhancing and boundary-finding mechanism is “reciprocal inhibition,” in which activity from one point in visual space inhibits activity from surrounding points and vice-versa. This internal competition tends to bring into focus the structure inherent in and afforded by the input (Hinton & Sejnowsky 1999).

**13. Supervised Learning.** This kind of unsupervised clustering based on enhancing structural similarities and correlations will not work,

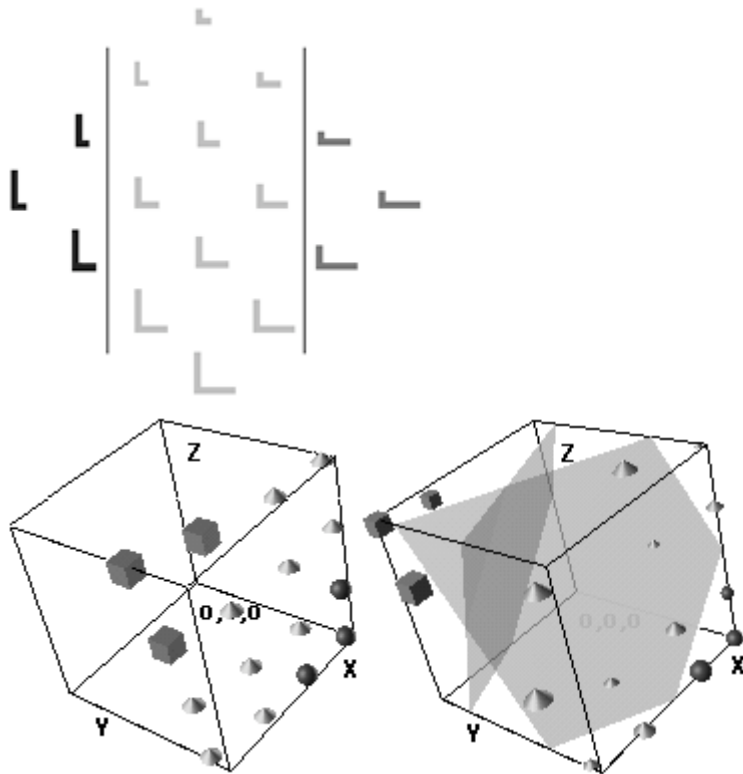


however, if different ways of clustering the very same sensory shadows are correct, depending on other circumstances (context-dependent categorization). To sort this out, supervision by error-corrective feedback is needed too; the sensorimotor structure and its affordances alone are not enough. We might say that supervised categories are even more underdetermined than unsupervised ones. Both kinds of category are underdetermined, because the sensory shadows of their members are made up of a high number of dimensions and features, their possible combinations yielding an infinity of potential shadows, making the subset of them that will afford correct categorization hard to find. But supervised categories have the further difficulty that there are many correct categorizations (sometimes an infinite number) for the very same set of shadows.

If you doubt this, open a dictionary again, pick any content word, say, “table,” then think of an actual table, and think of all the other things you could have called it (thing, object, vegetable, handiwork, furniture, hardwood, Biedermeyer, even “Charlie”). The other names you could have given it correspond to other ways you could have categorized it. Every category has both an “extension” (the set of things that are members of that category) and an “intension” (the features that make things members of that category rather than another). Not only are all things the members of an infinite number of different categories, but each of their features, and combinations of features is a potential basis (affordance) for assigning them to still more categories. So far, this is again just ontology. But if we return to sensory inputs, and the problem facing the theorist trying to explain how sensorimotor systems can do what they do, then sensory inputs are the shadows of a potentially infinite number of different kinds of things. Categorization is the problem of sorting them correctly, depending on the demands of the situation.

Supervised learning can help; if unsupervised learning (“mere exposure”) cannot find the winning features, perhaps feedback-guided trial and error training will do it, as with the pigeon’s black/white sorting and the chicken-sexing. There are some supervised learning algorithms so powerful that they are guaranteed to find the needle in the haystack, no matter how undetermined it is – as long as it is just underdetermined, not indeterminate (like the exact midpoint between black and white) or NP-complete -- and as long there is enough data and feedback and time (as, for the language-learning child, there is not, hence the “poverty of the stimulus”; Wexler 1991). Our categorization algorithms have to be able to do what we can do; so if we can categorize a set of inputs correctly, then those inputs must not only have the features that can afford correct categorization, but there must also be a way to find and use those affordances. (Figure 1 shows how a supervised neural net learns to sort a

set of forms into 3 categories by compressing and separating their internal representations in hidden unit space; Tijsseling & Harnad 1997.)



**Figure 1.** *Upper:* 3 sets of stimuli presented to neural net: vertical arm of L much longer, vertical and horizontal about equal, horizontal much longer. *Lower left:* Position of the hidden-unit representations of each of the three categories after auto-association but before learning (cubes represent Ls with long vertical arms, pyramids Ls with near-equal arms, spheres Ls with long horizontal arms). *Lower right:* Within-category compression and between-category separation when the net has learned to separate the three kinds of input. (From Tijsseling & Harnad 1997.)

**14. Vanishing Intersections?** Fodor and others have sometimes suggested otherwise: They have suggested that one of the reasons most categories can be neither learned nor evolved (and hence must be “innate” in some deeper sense than merely being a Darwinian adaptation) is the “vanishing intersections” problem: If you go back to the dictionary again, pick some content words, and then look for the “invariance” shared by all the sensory shadows of just about any of the things designated by those words, you will find there is none: their “intersection” is empty. What do all the shadows of boomerangs or tables – let alone Jerry Fodors or chicken-bottoms – have in common (even allowing dynamic

sensorimotor interactions with them)? And if that doesn't convince you, then what is the sensory shadow of categories like "goodness," "truth," or "beauty"?

**15. Direct Sensorimotor Invariants.** There is no reason for invariance theorists to back down from this challenge. First, it has to be pointed out that since we do manage to categorize correctly all those things designated by our dictionaries, there is indeed a capacity of ours that needs to be accounted for (see Appendix 1). To say that these categories are "innate" in a Cartesian, Platonic, or cosmogonic sense rather than just a Darwinian sense is simply to say that they are an unexplained, unexplainable mystery. So let us reject that. Let us assume that if organisms can categorize, then there must be a sensorimotor basis for that skill of theirs, and its source must be either evolution, learning, or both. Which means that there must be enough in those shadows to afford all of our categorization capacity.

**16. Abstraction and Hearsay.** Does it all have to be a matter of direct sensorimotor invariants, always? No, but the path to goodness, truth and beauty requires us to trace the chain of abstraction that takes us from categories acquired through direct sensory experience to those acquired through linguistic "hearsay":

Consider the five sensorimotor ways we can interact differentially with things, the five kinds of things we can do with things: We can see them, recognize them, manipulate them, name them or describe them. "Manipulate" in a sense already covers all five, because manipulating is something we do with things; but let us reserve the word "manipulate" for our more direct physical interactions with objects, such as touching, lifting, pushing, building, destroying, eating, mating with, and fleeing from them. Naming them and describing them is also a thing we do with them, but let us not subsume those two acts under manipulation. Seeing and recognizing are likewise things we do with things, but these too are better treated separately, rather than as forms of manipulation. And "seeing" is meant to stand in for all modes of sensory contact with things (hearing, smelling, tasting, touching), not just vision.

Recognizing is special, because it is not just a passive sensory event. When we recognize something, we see it as a kind of thing (or an individual) that we have seen before. And it is a small step from recognizing a thing as a kind or an individual to giving it a name. Seeing requires sensorimotor equipment, but recognizing requires more. It requires the capacity to abstract. To abstract is to single out some subset of the sensory input, and ignore the rest. For example, we may see many flowers in a scene, but we must abstract to recognize some of them as

being primroses. Of course, seeing them as flowers is itself abstraction. Even distinguishing figure from ground is abstraction. Is any sensorimotor event not abstraction?

**17. Abstraction and Amnesia.** To answer, we have to turn to fiction. Borges, in his 1944 short story, "Funes the Memorious," describes a person who cannot abstract. One day Funes fell off his horse, and from then onward he could no longer forget anything. He had an infinite rote memory. Every successive instant of his experience was stored forever; he could mentally replay the "tapes" of his daily experience afterwards, and it would take even longer to keep re-experiencing them than it had to experience them in the first place. His memory was so good that he gave proper names or descriptions to all the numbers -- "Luis Melián Lafinur, Olimar, azufre, los bastos, la ballena, el gas, la caldera, Napoléon, Agustín de Vedía" -- from 1 all the way up to enormous numbers (see Appendix 2). Each was a unique individual for him. But, as a consequence, he could not do arithmetic; could not even grasp the concepts counting and number. The same puzzlement accompanied his everyday perception. He could not understand why we people with ordinary, frail memories insist on calling a particular dog, at a particular moment, in a particular place, in a particular position, by the same name that we call it at another moment, a different time, place, position. For Funes, every instant was infinitely unique, and different instants were incomparable, incommensurable.

Funes's infinite rote memory was hence a handicap, not an advantage. He was unable to forget -- yet selective forgetting, or at least selective ignoring, is what is required in order to recognize and name things. Strictly speaking, a true Funes could not even exist, or if he did, he could only be a passive sensorimotor system, buffeted about by its surroundings (like the sand by the wind). Borges portrayed Funes as having difficulties in grasping abstractions, yet if he had really had the infinite memory and incapacity for selective forgetting that Borges ascribed to him, Funes should have been unable to speak at all, for our words all pick out categories bases on abstraction. He should not have been able to grasp the concept of a dog, let alone any particular dog, or anything else, whether an individual or a kind. He should have been unable to name numbers, even with proper names, for a numerosity (or a numeral shape) is itself an abstraction. There should be the same problem of recognizing either a numerosity or numeral as being the same numerosity (numeral) on another occasion as there was in recognizing a dog as the same dog, or as a dog at all.

**18. Invariance and Recurrence.** Funes was a fiction, but Luria described a real person who had handicaps that went in the same

direction, though not all the way to an infinite rote memory. In “The Mind of a Mnemonist” (1968) Luria describes a stage memory-artist, “S,” whom he had noticed when S was a journalist because he never took notes. S did not have an infinite rote memory like Funes’s, but a far more powerful and persistent rote memory than a normal person. When he performed as a memory artist he would memorize long strings of numbers heard only once, or all of the objects in the purse of an audience member. He could remember the exact details of scenes, or long sequences. He also had synaesthesia, which means that sensory events for him were richer, polysensory experiences: sounds and numbers had colors and smells; these would help him remember. But his powerful rote memory was a handicap too. He had trouble reading novels, because when a scene was described, he would visualize a corresponding scene he had once actually seen, and soon he was lost in reliving his vivid eidetic memory, unable to follow the content of the novel. And he had trouble with abstract concepts, such as numbers, or even ordinary generalizations that we all make with no difficulty.

What the stories of Funes and S show is that living in the world requires the capacity to detect recurrences, and that that in turn requires the capacity to forget or at least ignore what makes every instant infinitely unique, and hence incapable of exactly recurring. As noted earlier, Gibson’s (1979) concept of an “affordance” captures the requisite capacity nicely: Objects afford certain sensorimotor interactions with them: A chair affords sitting-upon; flowers afford sorting by color, or by species. These affordances are all invariant features of the sensory input, or of the sensorimotor interaction with the input, and the organism has to be capable of detecting these invariants selectively -- of abstracting them, ignoring the rest of the variation. If all sensorimotor features are somehow on a par, and every variation is infinitely unique, then there can be no abstraction of the invariants that allow us to recognize sameness, or similarity, or identity, whether of kinds or of individuals.

**19. Feature Selection and Weighting.** Watanabe’s (1985) “Ugly Duckling Theorem” captures the same insight. He describes how, considered only logically, there is no basis for saying that the “ugly duckling” -- the odd swanlet among the several ducklings in the Hans Christian Anderson fable -- can be said to be any less similar to any of the ducklings than the ducklings are to one another. The only reason it looks as if the ducklings are more similar to one another than to the swanlet is that our visual system “weights” certain features more heavily than others -- in other words, it is selective, it abstracts certain features as privileged. For if all features are given equal weight and there are, say, two ducklings and a swanlet, in the spatial position D1, S, D2, then although D1 and D2 do share the feature that they are both yellow, and S

is not, it is equally true that D1 and S share the feature that they are both to the left of D2 spatially, a feature they do not share with D2. Watanabe pointed out that if we made a list of all the (physical and logical) features of D1, D2, and S, and we did not preferentially weight any of the features relative to the others, then S would share exactly as many features with D1 as D1 shared with D2 (and as D2 shared with S). This is an exact analogue of Borges's and Luria's memory effect, for the feature list is in fact infinite (it includes either/or features too, as well as negative ones, such as "not bigger than a breadbox," not double, not triple, etc.), so unless some features are arbitrarily selected and given extra weight, everything is equally (and infinitely) similar to everything else.

But of course our sensorimotor systems do not give equal weight to all features; they do not even detect all features. And among the features they do detect, some (such as shape and color) are more salient than others (such as spatial position and number of feathers). And not only are detected features finite and differentially weighted, but our memory for them is even more finite: We can see, while they are present, far more features than we can remember afterward.

**20. Discrimination Versus Categorization.** The best illustration of this is the difference between relative and absolute discrimination that was pointed out by George Miller in his famous 1956 paper on our brains' information-processing limits: "The Magical Number 7 +/- 2". If you show someone an unfamiliar, random shape, and immediately afterward show either the same shape again or a slightly different shape, they will be able to tell you whether the two successive shapes were the same or different. That is a relative discrimination, based on a simultaneous or rapid successive pairwise comparison. But if instead one shows only one of the two shapes, in isolation, and asks which of the two it is, and if the difference between them is small enough, then the viewer will be unable to say which one it is. How small does the difference have to be? The "just-noticeable-difference" or JND is the smallest difference that we can detect in pairwise relative comparisons. But to identify a shape in isolation is to make an absolute discrimination (i.e., a categorization), and Miller showed that the limits on absolute discrimination were far narrower than those on relative discrimination.

Let us call relative discrimination "discrimination" and absolute discrimination "categorization." Differences have to be far greater in order to identify what kind or individual something is than for telling it apart from something else that is simultaneously present or viewed in rapid succession. Miller pointed out that if the differences are all along only one sensory dimension, such as size, then the number of JNDs we

can discriminate is very large, and the size of the JND is very small, and depends on the dimension in question. In contrast, the number of regions along the dimension for which we can categorize the object in isolation is approximately seven. If we try to subdivide any dimension more finely than that, categorization errors grow.

This limit on categorization capacity has its counterpart in memory too: If we are given a string of digits to remember we -- unlike Luria's S, who can remember a very large number of them -- can recall only about 7. If the string is longer, errors and interference grow.

**21. Recoding and Feature Selection.** Is there any way to increase our capacity to make categorizations? One way is to add more dimensions of variation; presumably this is one of the ways in which S's synaesthesia helped him. But even higher dimensionality has its limits, and never approaches the resolution power of the JND of sensory discrimination.

Another way of increasing memory is by recoding. Miller showed that if we have to remember a string of 0's and 1's, then a string of 7 items is about our limit. But if we first learn to recode the digits into, say, triplets in binary code, using their decimal names -- so that 001 is called "one", 010 is called "two," 011 is called "three" etc., and we overlearn that code, so that we can read the strings automatically in the new code, then we can remember three times as many of the digits. The 7-limit is still there, but it is now operating on the binary triplets into which we have recoded the digits: 101 is no longer three items: it is recoded into one "chunk," "five." We have learned to see the strings in terms of bigger chunks -- and it is these new chunks that are now subject to the 7-limit, not the single binary digits.

Recoding by overlearning bigger chunks is a way to enhance rote memory for sequences, but something similar operates at the level of features of objects: Although the number of features our sensory systems can detect in an object is not infinite, it is large enough so that if we see two different objects, sharing one or a few features, we will not necessarily be able to detect that they share features, hence that they are the same kind of object. This is again a symptom of the "underdetermination" mentioned earlier, and is related to the so-called "credit assignment problem" in machine learning: How to find the winning feature or rule among many possibilities (Sutton 1984)?

To be able to abstract the shared features, we need supervised categorization training (also called "reinforcement learning"), with trial and error and corrective feedback based on a large enough sample to allow our brains to solve the credit-assignment problem and abstract the

invariants underlying the variation. The result, if the learning is successful, is that the inputs are recoded, just as they are in the digit string memorization; the features are re-weighted. The objects that are of the same kind, because they share invariant features, are consequently seen as more similar to one another; and objects of different kinds, not sharing the invariants, are seen as more different.

This within-category enhancement of perceived similarity and between-category enhancement of perceived differences is again the categorical perception (CP) described earlier in the case of color. The sensory “shadows” of light frequency, intensity and saturation were recoded and re-weighted by our evolved color receptors so as to selectively detect and enhance the spectral ranges that we consequently see as red, yellow, etc.

## ***22. Learned Categorical Perception and the Whorf Hypothesis.***

When CP is an effect of learning, it is a kind of a Whorfian effect. Whorf (1956) suggested that how objects look to us depends on how we sort and name them. He cited colors as an example of how language and culture shape the way things look to us, but the evidence suggests that the qualitative color-boundaries along the visible spectrum are a result of inborn feature detectors rather than of learning to sort and name colors in particular ways. Learned CP effects do occur, but they are subtler than color CP, and can only be demonstrated in the psychophysical laboratory (Goldstone 1994, 2001; Livingston et al. 1998).

Figure 2 below illustrates this for a task in which subjects learned texture categorization. For an easy categorization task, there was no difference before and after learning, but for a hard one, learning caused within-category compression and between-category separation. (From Pevtzow & Harnad 1997).



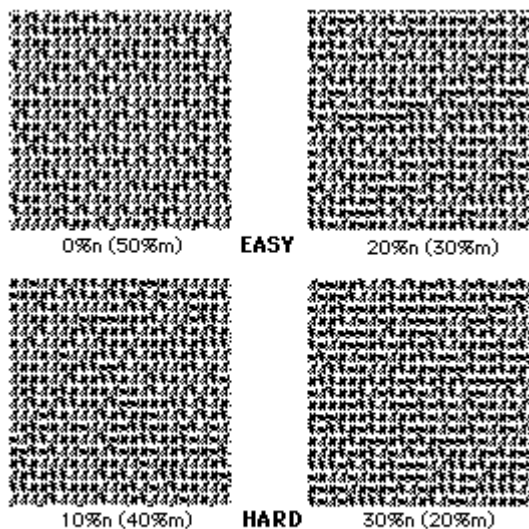


Figure 1. Examples of the Easy (upper) and Hard (lower) texture categories.

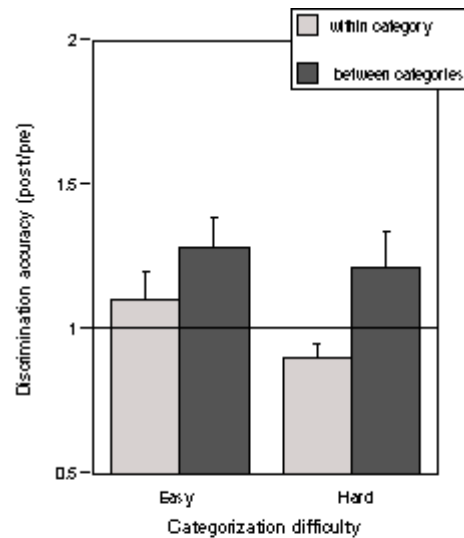


Figure 4. Ratios of discrimination accuracy after/before learning (Post/Pre) on the Easy and Hard task for Learners only. Separation is indicated by a ratio > 1 and compression by a ratio < 1. Error bars indicate standard error. There is significant compression within and near-significant separation between for the Hard Task but nonsignificant separation only for the Easy Task.

Figure 2. Left : Examples of the Easy (upper) and Hard (lower) texture categories. Right : Ratio of discrimination accuracy after/before learning (Post/Pre) in the Easy and the Hard task for Learners only. Separation is indicated by a ratio > 1 and compression by a ratio < 1. Error bars indicate standard error. There is a significant compression within and near-significant separation between for the Hard task but nonsignificant separation only for the Easy Task. (From Pevzow & Harnad 1997.)

Yet learned CP works much the way inborn CP does: Some features are selectively enhanced, others are suppressed, thereby bringing out the commonalities underlying categories or kinds. This works like a kind of input filter, siphoning out the categories on the basis of their invariant features, and ignoring or reducing the salience of non-invariant features. The supervised and unsupervised learning mechanisms discussed earlier have been proposed as the potential mechanisms for this abstracting capacity, with sensorimotor interactions also helping us to converge on the right affordances, resolving the underdetermination and solving the credit-assignment problem.

Where does this leave the concrete/abstract distinction and the vanishing-intersections problem, then? In what sense is a primrose

concrete and a prime number abstract? And how is “roundness” more abstract than “round,” and “property” more abstract still? Identifying any category is always based on abstraction, as the example of Funes shows us. To recognize a wall as a wall rather than, say, a floor, requires us to abstract some of its features, of which verticality, as opposed to horizontality, is a critical one here (and sensorimotor interactions and affordances obviously help narrow the options). But in the harder, more underdetermined cases like chicken–sexing, what determines which features are critical? (The gist of this underdetermination is there in the Maine joke: “How’s your wife?” “Compared to what?”)

**23. *Uncertainty Reduction.*** Although categorization is an absolute judgment, in that it is based on identifying an object in isolation, it is relative in another sense: What invariant features need to be selectively abstracted depend entirely on what the alternatives are, amongst which the isolated object needs to be sorted. “Compared to what?” The invariance is relative to the variance. Information, as we learn from formal information theory, is something that reduces the uncertainty among alternatives. So when we learn to categorize things, we are learning to sort the alternatives that might be confused with one another. Sorting walls from floors is rather trivial, because the affordance difference is so obvious already, but sorting the sex of newborn chicks is harder, and it is even rumoured that the invariant features are ineffable in that case: They cannot be described in words. That’s why the only way to learn them is through the months or years of trial and error reinforcement training guided by feedback under the supervision of masters.

**24. *Explicit Learning.*** But let us not mistake the fact that it is difficult to make them explicit verbally for the fact that there is anything invisible or mysterious about the features underlying chicken–sexing -- or any other subtle categorization. Biederman did a computer–analysis of newborn chick–abdomens and identified the winning invariants described in terms of his “geon” features (Biederman & Shiffrar 1987). He was then able to teach the features and rules through explicit instruction to a sample of novices so that within a short time they were able to sex chicks at the brown–belt level, if not the black belt level. This progress should have taken them months of supervised trial–and–error training, according to the grandmasters.

So if we accept that all categorization, great and small, depends on selectively abstracting some features and ignoring others, then all categories are abstract. Only Funes lives in the world of the concrete, and that is the world of mere passive experiential flow from one infinitely unique instant to the next (like the sand in the wind). For to do anything systematic or adaptive with the input would require abstraction, whether

innate or learned: the detection of the recurrence of a thing of the same kind.

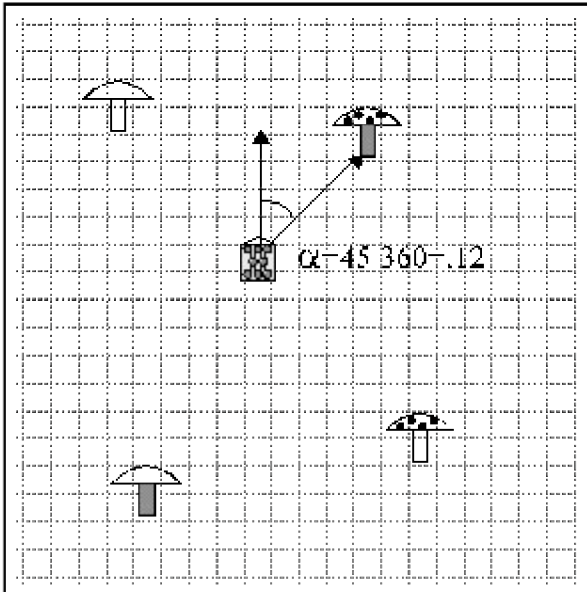
**25. *Categorization Is Abstraction.*** What about degrees of abstractness? (Having, with G.B. Shaw, identified categorization's profession – abstraction -- we are now merely haggling about the price.) When I am sorting things as instances of a round-thing and a non-round-thing, I am sorting things. This thing is round, that thing is non-round. When I am sorting things as instances of roundness and non-roundness, I am sorting features of things. Or rather, the things I am sorting are features (also known as properties, when we are not just speaking about them in a sensorimotor sense). And features themselves are things too: roundness is a feature, an apple is not (although any thing, even an apple, can also be a part, hence a feature, of another thing).

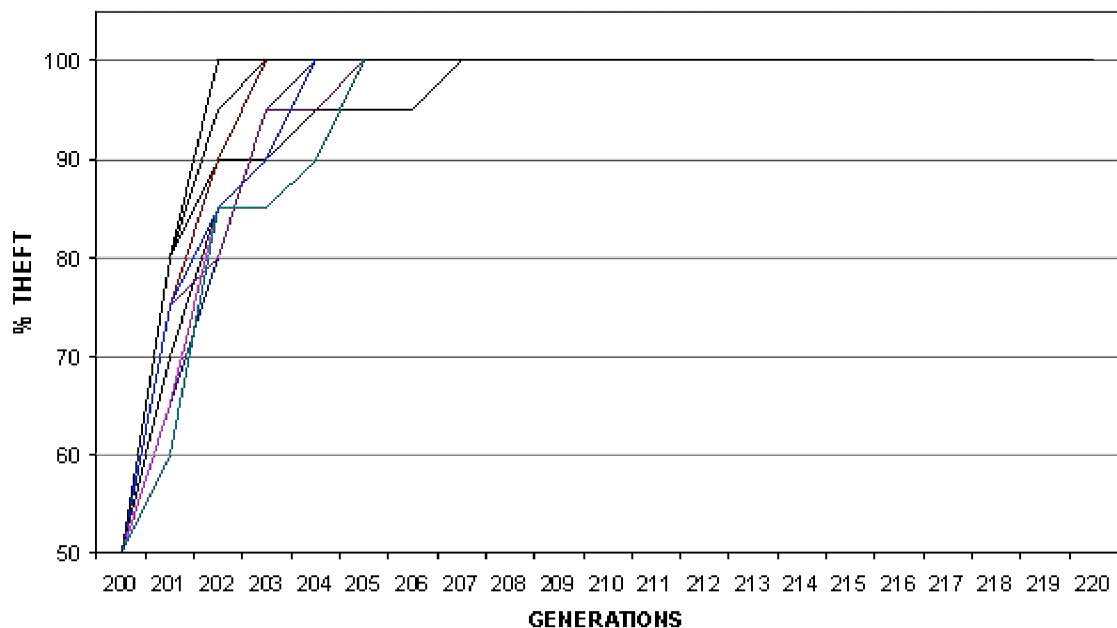
**26. *Sensorimotor Grounding: Direct and Derivative.*** In principle, all this sorting and naming could be applied directly to sensorimotor inputs; but much of the sorting and naming of what we consider more abstract things, such as numbers, is applied to symbols rather than to direct sensorimotor interactions with objects. I name or describe an object, and then I categorize it: "A number is an invariant numerosity" (ignoring the variation in the kinds or individuals involved). This simple proposition already illustrates the adaptive value of language: Language allows us to acquire new categories indirectly, through "hearsay," without having to go through the time-consuming and risky process of direct trial-and-error learning. Someone who already knows can just tell me the features of an X that will allow me to recognize it as an X.

(This is rather like what Biederman did for his experimental subjects, in telling them what features to use to sex chickens, except that his method was not pure hearsay, but hybrid: It was show-and-tell, not just tell, because he did not merely describe the critical features verbally; he also pointed them out and illustrated them visually. He did not first pretrain his subjects on geon-naming, as Miller's subjects were pretrained on naming binary triplets.)

**27. *The Adaptive Advantage of Language: Hearsay.*** If Biederman had done it all with words, through pure hearsay, he would have demonstrated the full and unique category-conveying power of language: In sensorimotor learning, the abstraction usually occurs implicitly. The neural net in the learner's brain does all the hard work, and the learner is merely the beneficiary of the outcome. The evidence for this is that people who are perfectly capable of sorting and naming things correctly usually cannot tell you how they do it. They may try to tell you what

features and rules they are using, but as often as not their explanation is incomplete, or even just plain wrong. This is what makes cognitive science a science; for if we could all make it explicit, merely by introspecting, how it is that we are able to do all that we can do, then our introspection would have done all of cognitive science's work for it (see Appendix 1). In practice we usually cannot make our implicit knowledge explicit, just as the master chicken-sexers could not. Yet what explicit knowledge we do have, we can convey to one another much more efficiently by hearsay than if we had to learn it all the hard way, through trial-and-error experience. This is what gave language the powerful adaptive advantage that it had for our species Cangelosi & Harnad 2001; see Figure 3).





**Figure 3.** An artificial-life simulation of mushroom foragers. Mushroom-categories could be learned in two different ways, by sensorimotor “toil” (trial-and-error learning with feedback from the consequences of errors) or linguistic “theft” (learning from overhearing the category described; hearsay). Within a very few generations the linguistics “thieves” out-survive and out-reproduce the sensorimotor toilers. (But note that the linguistically based categories must be grounded in sensorimotor categories: it cannot be theft all the way down.) (From Cangelosi & Harnad 2001.)

Where does this leave prime numbers then, relative to primroses? Pretty much on a par, really. I, for one, do not happen to know what primroses are. I am not even sure they are roses. But I am sure I could find out, either through direct trial and error experience, my guesses corrected by feedback from the masters, and my internal neural nets busily and implicitly solving the credit-assignment problem for me, converging eventually on the winning invariants; or, if the grandmasters are willing and able to make the invariants explicit for me in words, I could find out what primroses are through hearsay. It can't be hearsay all the way down, though. I will have had to learn some things ground-level things the hard, sensorimotor way, if the words used by the grandmasters are to have any sense for me. The words would have to name categories that I already have.

Is it any different with prime numbers? I know they are a kind of number. I will have to be told about factoring, and will probably have to try it out on some numbers to see what it affords, before recognizing that some kinds of numbers do afford factoring and others do not. The same is true for finding out what deductive proof affords, when they tell me more about further features of prime numbers. Numbers themselves I will have had to learn at first hand, supervised by feedback in absolutely discriminating numerosities, as provided by yellow-belt arithmeticians – for here too it cannot be hearsay all the way down. (I will also need to experience counting at first hand, and especially what “adding one” to something, over and over again, affords.)

**28. Absolute Discriminables and Affordances.** But is there any sense in which primroses or their features are “realer” than prime numbers and their features? Any more basis for doubting whether one is really “out there” than the other? The sense in which either of them is out there is that they are both absolute discriminables: Both have sensorimotor affordances that I can detect, either implicitly, through concrete trial-and-error experience, guided by corrective feedback (not necessarily from a live teacher, by the way: if, for example, primroses were edible, and all other flowers toxic, or prime numerosities were fungible, and all others worthless, feedback from the consequences of the sensorimotor interactions would be supervision enough); or explicitly, through verbal descriptions (as long as the words used are already grounded, directly or recursively, in concrete trial-and-error experience; Harnad 1990). The affordances are not imposed by me; they are “external” constraints, properties of the outside world, if you like, governing its sensorimotor interactions with me. And what I do know of the outside world is only through what it affords (to my senses, and to any sensory prostheses I can use to augment them). That  $2+2$  is 4 rather than 5 is hence as much of a sensorimotor constraint as that projections of nearer objects move faster along my retina than those of farther ones.

**29. Cognitive Science is Not Ontology.** Mere cognitive scientists (sensorimotor roboticists, really) should not presume to do ontology at all, or should at least restrict their ontic claims to their own variables and terms of art -- in this case, sensorimotor systems and their inputs and outputs. By this token, whatever it is that “subtends” absolute discriminations -- whatever distal objects, events or states are the sources of the proximal projections on our sensory surfaces that afford us our capacity to see, recognize, manipulate, name and describe them -- are all on an ontological par; and subtler discriminations are unaffordable.

Where does this leave goodness, truth and beauty, and their sensorimotor invariants? Like prime numbers, these categories are acquired largely by hearsay. The ethicists, jurists and theologians (not to mention our parents) tell us explicitly what kinds of acts and people are good and what kind are not, and why (but the words in their explicit descriptions must themselves be grounded, either directly, or recursively, in sensorimotor invariants: again, categories cannot be hearsay all the way down.). We can also taste what's good and what's not good directly with our senses, of course, in sampling some of their consequences. We perhaps rely more on our own sensory tastes in the case of beauty, rather than on hearsay from aestheticians or critics, though we are no doubt influenced by them and by their theories too. The categories "true" and "false" we sample amply through direct sensory experience, but there too, how we cognize them is influenced by hearsay; and of course the formal theory of truth looks more and more like the theory of prime numbers, with both constrained by the affordances of formal consistency.

**30. Cognition Is Categorization.** But, at bottom, all of our categories consist in ways we behave differently toward different kinds of things, whether it be the things we do or don't eat, mate with, or flee from, or the things that we describe, through our language, as prime numbers, affordances, absolute discriminables, or truths. And isn't that all that cognition is for -- and about?

## References

Berlin B & Kay P (1969) *Basic color terms: Their universality and evolution*. University of California Press, Berkley

Biederman, I. & Shiffrar, M. M. (1987) Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 13: 640 - 645. <http://www.phon.ucl.ac.uk/home/richardh/chicken.htm>

Borges, J.L. (1962) *Funes el memorioso*  
<http://www.bridgewater.edu/~atrupe/GEC101/Funes.html>

Cangelosi, A. & Harnad, S. (2001) The Adaptive Advantage of Symbolic Theft Over Sensorimotor Toil: Grounding Language in Perceptual Categories. *Evolution of Communication* 4(1) 117-142  
<http://cogprints.soton.ac.uk/documents/disk0/00/00/20/36/index.htm>

Catania, A.C. & Harnad, S. (eds.) (1988) *The Selection of Behavior. The Operant Behaviorism of BF Skinner: Comments and Consequences*. New York: Cambridge University Press.

Chomsky, N. (1976) In Harnad, Stevan and Steklis, Horst D. and Lancaster, Jane B., Eds. *Origins and Evolution of Language and Speech*, [page 58](#). *Annals of the New York Academy of Sciences*.

Fodor, J. A. (1975) [The language of thought](#). New York: Thomas Y. Crowell

Fodor, J. A. (1981) *RePresentations*. Cambridge MA: MIT/Bradford.

Fodor, J. A. (1998). *In critical condition: Polemical essays on cognitive science and the philosophy of mind*. Cambridge, MA: MIT Press.

Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston. (Currently published by Lawrence Erlbaum, Hillsdale, NJ <http://cognet.mit.edu/MITECS/Entry/gibson1>

Goldstone, R.L., (1994) [Influences of categorization on perceptual discrimination](#). *Journal of Experimental Psychology: General* 123: 178–200

Goldstone, R.L. (2001) The Sensitization and Differentiation of Dimensions During Category Learning. *Journal of Experimental Psychology: General* 130: 116–139

Harnad, S. (1987) Category Induction and Representation, In: Harnad, S. (ed.) (1987) *Categorical Perception: The Groundwork of Cognition* . New York: Cambridge University Press.  
<http://cogprints.soton.ac.uk/documents/disk0/00/00/15/72/index.html>

Harnad, S. (1990) The Symbol Grounding Problem. *Physica D* 42: 335–346.  
<http://cogprints.soton.ac.uk/documents/disk0/00/00/06/15/index.html>

Harnad, S. (2000) Minds, Machines, and Turing: The Indistinguishability of Indistinguishables. *Journal of Logic, Language, and Information* 9(4): 425–445. (special issue on "Alan Turing and Artificial Intelligence")  
<http://cogprints.soton.ac.uk/documents/disk0/00/00/16/16/index.html>

Harnad, S. (2001) No Easy Way Out. *The Sciences* 41(2) 36–42.  
<http://cogprints.soton.ac.uk/documents/disk0/00/00/16/23/index.html>



Harnad, S. (2003) Categorical Perception. *Encyclopedia of Cognitive Science*. Nature Publishing Group. Macmillan.  
<http://www.ecs.soton.ac.uk/~harnad/Temp/catperc.html>

Harnad, S. (2003) Symbol–Grounding Problem. *Encyclopedia of Cognitive Science*. Nature Publishing Group. Macmillan.  
<http://www.ecs.soton.ac.uk/~harnad/Temp/symgro.htm>

Hinton, G. & Sejnowsky, T. (eds.) (1999) *Unsupervised Learning : Foundations of Neural Computation*. MIT Press.

Livingston, Kenneth and Andrews, Janet and Harnad, Stevan (1998) Categorical Perception Effects Induced by Category Learning. *Journal of Experimental Psychology: Learning, Memory and Cognition* 24(3):732–753 <http://eprints.ecs.soton.ac.uk/archive/00006883/>

Luria, A. R. (1968) *The Mind of a Mnemonist*. Harvard University Press  
<http://qsilver.queensu.ca/~phil158a/memory/luria.htm>

Miller, George (1956) The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review* 63:81–97 <http://cogprints.ecs.soton.ac.uk/archive/00000730/>

Pevtzow, R. & Harnad, S. (1997) Warping Similarity Space in Category Learning by Human Subjects: The Role of Task Difficulty. In: Ramsar, M., Hahn, U., Cambouropoulos, E. & Pain, H. (Eds.) *Proceedings of SimCat 1997: Interdisciplinary Workshop on Similarity and Categorization*. Department of Artificial Intelligence, Edinburgh University: 189 – 195.  
<http://cogprints.soton.ac.uk/documents/disk0/00/00/16/07/index.html>

Rosch, E. & Lloyd, B. B. (1978) *Cognition and categorization*. Hillsdale NJ: Erlbaum Associates

Sutton, R. S. (1984). *Temporal Credit Assignment in Reinforcement Learning*. PhD thesis, Department of Computer and Information Science, University of Massachusetts.

Steklis, Horst Dieter and Harnad, Stevan (1976) From hand to mouth: Some critical stages in the evolution of language, In: *Origins and Evolution of Language and Speech* (Harnad, Stevan, Steklis, Horst Dieter and Lancaster, Jane B., Eds.), 445–455. *Annals of the New York Academy of Sciences* 280.  
<http://cogprints.soton.ac.uk/documents/disk0/00/00/08/66/index.html>

Tijsseling, A. & Harnad, S. (1997) Warping Similarity Space in Category Learning by Backprop Nets. In: Ramscar, M., Hahn, U., Cambouropoulos, E. & Pain, H. (Eds.) Proceedings of SimCat 1997: Interdisciplinary Workshop on Similarity and Categorization. Department of Artificial Intelligence, Edinburgh University: 263 – 269.  
<http://cogprints.soton.ac.uk/documents/disk0/00/00/16/08/index.html>

Watanabe, S., (1985) "Theorem of the Ugly Duckling", *Pattern Recognition: Human and Mechanical*. Wiley  
<http://www.kamalnigam.com/papers/thesis-nigam.pdf>

Wexler, K. (1991) The argument from poverty of the stimulus. In : A. Kasher (ed.) *The Chomskyan Turn*. Cambridge. Blackwell.

Whorf, B.L. (1956) *Language, Thought and Reality*. (J.B. Carroll, Ed.) Cambridge: MIT <http://www.mtsu.edu/~dlavery/Whorf/blwquotes.html>

### **Appendix 1. There is nothing wrong with the “classical theory” of categorization.**

Eleanor Rosch has suggested that because we cannot state the invariant basis on which we categorize, that invariance must not exist (Rosch & Lloyd 1978), and hence there is something wrong with the so-called “classical” theory of categorization, according to which we categorize on the basis of the invariant features that are necessary and sufficient to afford categorization.

Not only do I not think there’s anything the least bit wrong with that “classical theory,” but I am pretty confident that there is no non-magic alternative to it. Rosch’s alternative was to vacillate rather vaguely between the idea that we categorize on the basis of “prototypes” or on the basis of “family resemblances”. Let’s consider each of these candidate mechanisms in turn:

To categorize on the basis of prototypes would be to identify a bird as a bird because it looks more like the template for a typical bird than the template for a typical fish. This would be fine if all, many, or most of the things we categorize indeed had templates, and our internal categorization mechanism could sort their sensory shadows by seeing which template they are closest to; in other words, it would be fine if such a mechanism could actually generate our categorization capacity.

Unfortunately it cannot. Template-matching is not very successful among the many candidate machine-learning models, and one of the reasons it is unsuccessful is that it is simply not the case that everything is a member of every category, to different degrees. It is not true ontologically that a bird is a fish (or a table) to a certain degree; nor is it true functionally that sensory shadows of birds can be sorted on the basis of their degree of similarity to prototypical birds, fish or tables. So prototype/template theory is a non-starter as a mechanism for our categorization capacity. It might explain our typicality judgments – is this a more typical bird than that -- but being able to make a typicality judgment presupposes being able to categorize; it does not explain it: Before I can say how typical a bird this is, I first need to identify it as a bird!

So if not prototypes, what about family-resemblances, then? What are family resemblances? They are merely a cluster of either/or features: This is an X, if it has feature A or B or not C. Either/or features (disjunctive invariants) are perfectly classical (so forget about thinking of family-resemblances as alternatives to classical theories of categorization). The problem is that saying that some features are either/or features leaves us no closer to answering “how” than we were before we were informed of this. Yes, some of the affordances of sensory shadows will be either/or features, but what we need to know is what mechanism will be able to find them!

The last Roschian legacy to category theory is the “basic object” level, vs. the superordinate or subordinate level. Here too it is difficult to see what, if anything, we have learned from Roschian theory.

If you point to an object, say, a table, and ask me what it is, chances are that I will say it's a table, rather than a Biedermeyer, or furniture, or “Charlie”. So what? As mentioned earlier, there are many ways to categorize the same objects, depending on context. A context is simply a set of alternatives among which the object's name is meant to resolve the uncertainty (in perfectly classical information-theoretic terms). So when you point to a table and ask me what it is, I pick “table” as the uncertainty-resolver in the default context (I may imagine that the room contains one chair, one computer, one waste-basket and one table. If I imagine that it contains four tables, I might have to identify this one as the Biedermeyer; and if there are four Biedermeyers, I may have to hope you know I've dubbed this one “Charlie.”

So much for subordinate levels. The same path can be taken for superordinate levels. It all devolves on the old Maine joke, which comes close to revealing a profound truth about categories: “How's Your wife?”

Reply: “Compared to what?” If we were just discussing the relative amount you should invest in furniture in your new apartment, as opposed to accessories, and you forgot what was in the adjacent room and asked what was in there (when there was just a table) I might reply furniture. If we were discussing ontology, I might say “vegetable” (as opposed to animal or mineral). Etc.

So citing the “basic object level” does not help explain our categorization capacity; that’s just what one arbitrarily assumes the default context of interconfusable alternatives to be, given no further information. The only sense in which “concrete” objects, directly accessible to our senses, are somehow more basic, insofar as categorization is concerned, than more “abstract” objects, such as goodness, truth or beauty is that sensorimotor categories must be grounded in sensory experience and the content of much of that experience is fairly similar and predictable for most members of our species.

### **Appendix Two. Associationism begs the question of categorization.**

The problem of association is the problem of rote pairing: an object with an object, a name with a name, a name with an object. Categorization is the problem of recognizing and sorting objects as kinds based on finding the invariants underlying sensorimotor interactions with their shadows. Associationism had suggested that this was just a matter of learning to associate tokens (instances, shadows) of an object-type with tokens of its type-name -- as indeed it is, if only we can first figure out which object-tokens are tokens (shadows) of the same object-type! Which is in turn the problem of categorization. Associationism simply bypassed the real problem, and reduced learning to the trivial process of rote association, governed by how often two tokens co-occurred (plus an unexplicated influence of how “similar” they were to one another).

Some associative factors are used by contemporary unsupervised learning models, where internal co-occurrence frequencies and similarities are used to cluster inputs into kinds by following and enhancing the natural landscape of their similarities and dissimilarities. But this is internal association among representational elements and patterns (e.g., units in a neural network), not external association between input tokens. And its scope is limited, as we have seen, for most of the shadows of most of the members of most of the categories in our dictionary could not be sorted into their respective categories by unsupervised association alone, because of underdetermination. Nor is supervised learning merely rote association with the added associative cue of the category-name (as provided by the supervisory feedback). The hard work in these learning models is done by the algorithm that solves the credit-assignment problem by finding the winning invariance in the haystack -- and no

model can do this at human categorization–capacity scale just yet. Critics of associationism, however, drew the incorrect conclusion that because (1) we don't know what the invariance is in most cases and (2) association is ill–equipped to find it in any case, it follows that there either is no invariance, or our brains must already know it innately in some mysterious way.