

## Hippocampal Theories and the Information/Computation Perspective

William B Levy

### INTRODUCTION

Through the years, there seems to be no end of theories about the function of the hippocampus. Such functions include olfactory processing, recent memory, memory consolidation, response inhibition, selective attention (see Douglas 1967 for review), spatial mapping (O'Keefe and Nadel 1978), stimulus mismatch operations, temporal and sequence learning (Schmajuk 1987), storage of declarative knowledge, and working memory. Squire (1982) reviewed many of the psychological theories of hippocampal function. Worse still, the functioning rat's hippocampus may not do what H.M.'s hippocampus does (compare Scoville and Milner 1957 to above citations).

The point is that, after 30 plus years of serious experimental and intellectual assaults, we have a theory of hippocampal function that is only a little better than when we started. Certainly none of these theories is very sharp nor is there any particular trend for such theories becoming precise statements of hippocampal function. Even more bothersome is the fact that none of these theories relates function to the rather special anatomy of the hippocampus or to what we have learned or even to what we will learn about the cellular physiology of the hippocampus. Clearly this old approach to understanding hippocampal function alone will not do.

As a reductionistic neuroscientist, I would like to understand how groups of cells function to form an integrated system. This approach generally tries to relate cellular physiology and anatomy to system physiology. Within the context of hippocampal function, such an approach seeks to understand the input-output characteristics of the hippocampus as a system from its parts and to relate this system's behavior to the behavior of other brain regions and to brain functions. If such an abstract approach is properly framed, certain questions will be natural, while other questions will be easily avoided. Consider the four examples that follow.

First, we should not have to know the function of the entorhinal cortex (EC), the major extrinsic input to the hippocampus, or the Hubel and Wiesel-like feature-detecting properties of entorhinal neurons before we can study the hippocampus. Otherwise, by a natural regression, we would have to know the feature codings of all the brain regions that precede the hippocampus in sensory processing before we formulate a theory of hippocampal function. This approach certainly would postpone research on the hippocampus for a long time.

Second, we should be able to compare input-output functions between areas and between species. For example, how do primary visual cortex (area 17) and the hippocampus differ in the way they process signals? How do the rat and human hippocampi differ in their signal-processing abilities?

Third, a proper framework should stimulate quantitative as well as qualitative statements of function. Thus, when there are thousands of stimulating and recording electrodes in the future, there will be predictions worth testing.

Finally, independently obtained results from various brain regions should easily fit together. That is, if a proper language of input-output functions is used consistently for different regions, then we should be able to plug various parts of the brain together after studying individual systems and obtain a prediction of total brain function.

Therefore, the idea underlying the proposed approach is to study the abstract input-output transformations of systems like the hippocampus at the level of cells and groups of cells over physiologically relevant ranges of parameters. Even this is not enough of a framework, however, to gain an understanding of the function of the hippocampus or of any other brain region. We have to include, in our framework, questions of some central but general issue of brain function itself. The issue should guarantee relevance to the entire organism and must be identifiable in a consistent way in isolated brain regions.

But what is it that is special about the brain's function?

The function of the brain is to control the environment and the organism so that the organism survives and produces fertile offspring, thereby ensuring the survival of the species. This control is accomplished in large part by the brain's abilities to *predict* (1) future states of the environment, (2) how current and future states of the environment will affect the organism, and (3) how the organism's interactions with the environment will alter the environment.

Central to all of this predicting is our ability to find, store, and use the predictiveness—the regularities—that exist in the environment. Successful

prediction is possible, because there are spatiotemporal regularities and physical constraints in the environment. We will refer to these regularities and constraints as "redundancies" for four reasons: (1) the formal mathematical measures suggested by this term; (2) the ease of identifying this formalized redundancy with what we call correlations, concepts, associations, regularities, and constraints; (3) the formalization applies both to the environment and to individual brain regions; and (4) the meaning of the word redundancy, at least at its extremes in these mathematical formulations, agrees with the meaning of the word in conversation. Thus, successful prediction of the environment depends upon environmental redundancy.

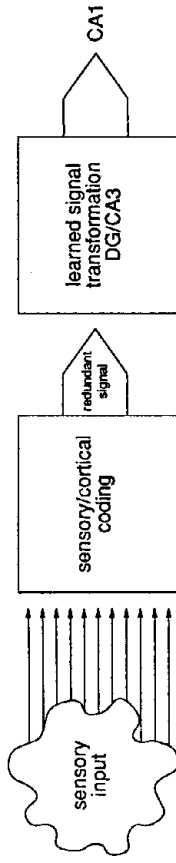
Stored redundancies can be used to form predictions. When a new signal arrives, it is transformed to a less redundant signal based on past associations. These transformed signals are called predictive representations and are synonymous with recodings and with coded representations and are, in a mathematical sense, predictions.

It is, then, the storage of redundancy at synapses and predictive representation by reduction of signal redundancy that finally frames the abstract approach to understanding brain function. This issue gives our abstract approach to the brain and its regions that extra something that takes it beyond simple input-output questions to a general problem central to brain function that is applicable to almost any brain region as well as to the environment itself.

#### APPLICATION OF THE APPROACH TO THE HIPPOCAMPUS

As a general issue, the reduction of redundancy is particularly easy to apply when there is only one class of high-information-carrying inputs. As an example, let's consider the dentate gyrus (DG)-CA3 system of the hippocampus.

Given this approach, the theory of DG-CA3 function is, loosely stated, as follows. The layer-II stellate cells of the EC provide information to the neurons of the DG-CA3 system. These neurons then reduce the signal redundancy while maintaining a distinctiveness of coding that allows for minimal loss of information. That is, the DG-CA3 system recodes the EC afferent information before sending it to the hippocampal CA1 region. The recoding is based on synaptically encoded associations, i.e., environmental redundancies, that have been learned via associative synaptic modification. This recoding is helpful to CA1, or to any other brain area, because of the severely restricted size and processing speed of CA1 (and of every other brain region for that matter; see below). Figure 1 schematically illustrates these simplifications.



**FIGURE 1.** *Recoding a signal to remove redundancy*

Note that there is only one class of inputs in figure 1. This simplification comes from a basic idea in information theory (Shannon and Weaver 1949) that information capacity increases with both the number of channels and the maximum frequency of each channel. In this view, the EC afferents to the DG-CA3 system carry essentially all of the information to this part of the hippocampus because of their large numbers and their high-frequency firing rates. This is not to say that the monoaminergic and cholinergic afferents are unimportant—indeed, I am as certain as anyone that the hippocampus would not function properly without them—but that their function is not to carry high-density information to the hippocampus. This job is left for entorhinal afferents. Note that this view does not take away from the importance of monoaminergic or cholinergic inputs. To gain a better understanding, consider a metaphor of a TV set with two wires going in: the power cord, identified with the nonspecific inputs, and the antenna wire, identified with the entorhinal afferents. The antenna wire carries the information, but cut the power cord and nothing works.

Since there is no other high-density, extrahippocampal information source interacting with the entorhinal input, the EC signal is interacting only with itself: but to what end? Certainly the EC is the beginning of the funnel that brings together signals from all the diverse sensory systems. The DG-CA3 helps in this funneling, so obviously a recoding is going on, but why? Action potentials are sure, so there is no reason to insert redundancy. Textbooks to the contrary, there are no relay areas that are not processors. To merely pass on a signal not only wastes time and processing units, but it risks the almost certain injection of noise into the signals. On the other hand, reducing signal redundancy will help CA1, whatever its task. By the way, CA1 desperately needs help, because the number of cells in CA1 is woefully inadequate to create, much less compare, all possible hypotheses.

Now, note that we concern ourselves not with the environment that the entire organism experiences but with the environment that the hippocampus itself experiences. This is a major simplification of the framework compared to the old approach. That is, aside from hormonal influences, the environment of the hippocampus is no more and no less than the information provided by its afferents.

Thus, we no longer must identify cells by their feature-detection properties relative to the environment external to the organism. Instead, we determine the redundancy reduction at each stage of processing, though we always have the option of speculating on the redundancy of the environment and the feature-detection codings that are, in fact, predictive representations.

It is not that environmental redundancy is an uninteresting question; it is just that it requires solving a more difficult problem than understanding the brain itself, i.e., understanding the redundancy or organization of the entire world. Also, it implicitly evokes an almost dogmatic belief that we know what we see and that what we consciously see are the proper dimensions of the world. I am suspicious that we no more know the component features of what we see than we know the component features of what we smell.

### A LITTLE MORE ABOUT REDUNDANCY

The regularities of the world are, through learning and heredity, stored in the connections—indeed as the quantitative connections—between neurons. Thus, we should understand that the associations that are learned and stored at synapses are synonymous with the regularities, correlations, and constraints, i.e., the redundancies, of the environment. Information and coding theories allow us to abstract and quantify the predictability of the environment using the idea of redundancy.

More precisely, redundancy is the amount of statistical dependence in a group of signals. When a linear measure of redundancy is desired, the mathematical formulations are quite explicit (Watanabe 1969).

The importance of predictive reduction of redundancy deserves comment. The reasons are both practical and philosophical. Practically, the brain is too small and too slow. This may seem strange, since we have always tried to impress the lay public and our students with the notion of how big the brain is—more than  $10^{10}$  cells, more than  $10^{14}$  synapses—but consider how many possible games of chess there are, and the brain is small in comparison. Consider how many possible configurations of the environment there are, each one you might encounter and need to represent, and the brain is minuscule by comparison.

Because of the overwhelming numbers of possible patterns in the world, it is critical to remove as much redundancy as possible from the representations of the environment. In addition, reducing redundancies will speed up computations and decrease storage requirements. Furthermore, when redundancies are removed from representations, there are some safe, fast, and easy ways to process representations and make predictions. Even these improvements will not be enough, however, and neural computations will usually have to “cheat” to compute fast enough for real time interactions. By cheating I mean make predictions, or equivalently do constraint

satisfaction, without examining each individually possible configuration of the environment. Such incomplete examination risks error. By removing redundancies, the average size of this error can be minimized.

The final argument may appeal only to the more philosophically minded but has application far beyond the hippocampus. First, note that the configurations of the environment that reach our sensory receptors are in fact signals. Then, the redundancy in these special signals is all there is to learn about the environment. In signals with no redundancy or in an environment with no redundancy, there is nothing to learn. Obviously, then, if we can reduce the environment by removing and storing the redundancies (or use the predictive redundancies we have inherited) while maintaining pattern distinctiveness, we understand all there is to know about these signals with regard to themselves. Thus, identifying and storing redundancy is what learning and, therefore, much of synaptic modification is all about.

### AN EXAMPLE OF REDUNDANCY IN OUR DAY-TO-DAY LIFE

Imagine this scene—you are at home, reading a book, waiting to be picked up and driven to the airport. A horn sounds from outside. Honk! You get up and go to the door to leave. Note how the honk predicts the car outside. Now suppose we recreate this scene again except you are not reading but sitting at the window. Just at the same time the horn sounds, you see the arrival of the expected car. Note that the sound alone was enough to predict the car so that, in this second case, hearing the horn and seeing the car are to a large extent redundant in the information that your ride has arrived.

### AN EXAMPLE OF NEURAL REDUNDANCY

In figure 2, we consider a very simple nervous system with two afferent lines of signals.

An event is an active cell; PROB ( ) stands for probability of.

(A) There is zero redundancy if there is total independence; e.g.,

$$\text{PROB (EVENT 1 AND EVENT 2)} = \text{PROB (EVENT 1)} \cdot \text{PROB (EVENT 2)}$$

(B) There is redundancy if there is not statistical independence, e.g.,

$$\text{PROB (EVENT 1 AND EVENT 2)} \neq \text{PROB (EVENT 1)} \cdot \text{PROB (EVENT 2)}$$

FIGURE 2. Examples of neural redundancy

If the probability of joint activity, i.e., the probability of both cells being active, is equal to the product of the individual probabilities of cell activity, as in (A) in figure 2, then there is statistical independence and no redundancy in the signals. If the equality does not hold, as in (B) in figure 2, then there is redundancy in the system. When there is no redundancy, conditional prediction is a waste of time, because there are no useful correlations or associations. When there is redundancy, we have something to gain from predictions about activity at one cell conditional on the activity of the other cell.

This idea of redundancy vs. no redundancy can be refined to formal measurements quantifying gradations that vary between no redundancy and total redundancy. One such measurement is an entropy called mutual information.

Using mutual information, we now state the theory more precisely:

The object of the DG-CA3 transformation is to minimize the mutual information of the CA3 output signals with respect to themselves, while maximizing the mutual information between EC input signals and CA3 output signals.

What we are proposing is that subsystems of the brain like the hippocampus face a dual optimization problem that involves (1) minimizing the mutual information of CA3 cell activities with regard to themselves, while, at the same time, (2) maximizing the mutual information of the layer-II EC input relative to the CA3 output. Part (1) is the optimization that reduces signal redundancy. The transformation performed by the dentate gyrus and CA3 recodes the EC representation of whatever into a CA3 representation of this same whatever. Notably the CA3 representation is less internally redundant than the EC representation. Part (2) is the optimization that maintains pattern distinctiveness. That is, those patterns that are easily separable or distinguishable in the EC representation are still easily distinguishable by their CA3 representations.

Just to show that this dual optimization problem is well formulated, we point out that there actually is a unique solution to this problem if the world is Gaussian (of course it is not!). For zero mean Gaussian environments, the Karhunen-Loeve transformation is the optimal mapping that uniquely solves this dual optimization (Watanabe 1969). This transformation codes the environment so that all output lines considered pairwise are independent and no higher order interactions exist. Because the transformation is invertible, the input-output mutual information is maximal. A more general solution suitable for optimally transforming more complex environments is desirable.

## CONCLUSION

Finally, we should realize that this new framework can happily exist with the old game. The following are examples of this coexistence.

1. Spatial sequences, such as encountered by small animals running through the maze of the underbrush and through drain pipes, require bringing together codings of rather diverse spatial relations and sequences. These spatial and spatiotemporal patterns are likely represented differently by the different senses (e.g., visual and olfactory codings must be very different initially). Recoding for an efficient compatible melding of these diverse dimensions seems essential to using all the cues quickly and accurately. Recoding to achieve redundancy reduction in DG-CA3 and other limbic structures may be critical to the degree of success or failure in such an environment.
2. Stimulus mismatch is a trigger for selective attention and an attempt to recode our stored, predictive redundancies. Mismatch activates certain nonspecific inputs to the hippocampus that in turn facilitate associative synaptic modification in the DG-CA3. This facilitation of modification in a mismatch situation is an opportunity to store redundancies that more accurately coincide with those of the environment.
3. Short-term memory incorporates a redundancy reduction process. This process saves space later when memories are transferred from short-term to long-term storage.

## HISTORICAL NOTES

The ideas expressed here have appeared at various times in various places. Included below are the authors who have been most influential on me or who, discovered in retrospect, presented relatively clear expositions of their ideas.

The ideas here are strongly related to those of S. Watanabe (1969) relating pattern recognition and inference to reduction of entropy.

The theme of redundancy reduction was quite explicitly introduced to neuroscience by Barlow (1959) and is still occasionally found as an issue in the visual system.

W.R. Ashby (1956) and Richard Bellman (1961) clearly expressed the idea that the nervous system or any other device was too small and too slow compared to the possibilities of the environment.

Marr (1970) mentions the issue of redundancy reduction in terms of the overwhelming encoding problems faced by the brain.

The unsupervised concept formation systems of various authors, including Anderson (1968), Amari (1977), Kohonen (1984), and Marr (1970) strongly resemble the problem and the system considered here. However, predictive reduction of redundancy is a more general issue than concept formation as it is usually presented. That is, the redundancy issue includes concept formation (as Marr points out). However, predictive redundancy reduction is less restrictive, and the mathematics are much better developed.

The equivalency between total understanding and recoding for minimum redundancy is almost a direct quote from the work by Rissanen (1986), while the germ of the idea appears in the writings of several philosophers.

J.A. Gray (1982) has emphasized the importance of prediction and related it to the limbic system.

The ideas of this chapter appeared previously in abstract form (Levy 1985).

## REFERENCES

- Amari, S.-I. Neural theory of association and concept-formation. *Biol Cybern* 26:175-185, 1977.
- Anderson, T.A. A memory storage model utilizing spatial correlation functions. *Kybernetik* 5:113-119, 1968.
- Ashby, W.R. Design for an intelligence-amplifier. In: Shannon, C.E., and McCarthy, J., eds. *Automata Studies*. Princeton: Princeton University Press, 1956. pp. 215-234.
- Barlow, H.B. Sensory mechanisms, the reduction of redundancy, and intelligence. In: *Mechanisation of Thought Processes*. London: Her Majesty's Stationery Office, 1959. pp. 537-559.
- Bellman, R. *Adaptive Control Processes: A Guided Tour*. Princeton: Princeton University Press, 1961. 255 pp.
- Douglas, R.J. The hippocampus and behavior. *Psychol Bull* 67:416-442, 1967.
- Gray, J.A. *The Neuropsychology of Anxiety: An Enquiry Into the Functions of the Septo-Hippocampal System*. New York: Oxford University Press, 1982. 548 pp.
- Kohonen, T. Correlation matrix memories. *IEEE Trans Computers* C-21:353-359, 1972.
- Kohonen, T. *Self-Organization and Associative Memory*. Berlin: Springer-Verlag, 1984. 255 pp.
- Levy, W.B. An information/computation theory of hippocampal function. *Soc Neurosci Abs* 11:493, 1985.
- Marr, D. A theory for cerebral neocortex. *Proc R Soc Lond B* 176:161-234, 1970.
- O'Keefe, J., and Nadel, L. *The Hippocampus as a Cognitive Map*. Oxford: Oxford University Press, 1978. 570 pp.

Rissanen, J. Stochastic complexity and modeling. *Ann Stat* 14:1080-1100, 1986.

Schmajuk, N.A. The hippocampus and the control of information storage in the brain. In: Arbib, M., and Amari, S.I., eds. *Competition and Cooperation in Neural Nets*. New York: Springer, 1987.

Scoville, W.B., and Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J Neurosurg Psychiat* 20:11-21, 1957.

Shannon, C.E., and Weaver, W. *The Mathematical Theory of Communication*. Chicago: University of Illinois Press, 1949. 125 pp.

Squire, L. The neuropsychology of human memory. *Ann Rev Neurosci* 5:241-273, 1982.

Watanabe, S. *Knowing and Guessing. A Quantitative Study of Inference and Information*. New York: John Wiley, 1969. 592 pp.

## ACKNOWLEDGMENTS

This work was supported by the U.S. Public Health Service grant NS 15488 from the National Institutes of Health and the RSDA grant MH 00622 from the National Institute of Mental Health to William B Levy.

## AUTHOR

William B Levy, Ph.D  
Research Associate Professor  
Department of Neurosurgery  
University of Virginia Health Sciences Center  
Box 420  
Charlottesville, VA 22908