6 How do brains detect novelty?

Endel Tulving

“We have to live today by what truth we can get today, and be ready tomorrow to call it falsehood.”

William James

Festschrifts are among the wonderful inventions of enlightened minds. They serve many purposes. The main purpose, of course, is to honour the honouree and make him happy. Sometimes this purpose does not work because not all great scholars and scientists for whom Festschrifts are written know how to be happy. In the present case, however, the success of this part of the venture is assured because our honouree, Lars-Göran Nilsson, not only believes in happiness but also practises it wherever and whenever possible. In that sense alone, this Festschrift in his honour is especially appropriate. Festschrifts also make Festschrift organizers happy, not only because they can thereby publicly express their respect and admiration for the honouree but also because it makes for a handsome addition to their own curricula vitae. And Festschrifts make the invited contributors happy because it sometimes affords them a chance to publish something that they might not be allowed to publish under less friendly circumstances. As a contributor I am grateful to Lars-Göran Nilsson for laying the groundwork for such an opening, and to Lars Bäckman and Lars Nyberg for effecting it, thereby making it possible to tell the story that appears here.

The story is about novelty. As everyone knows, brains are very good at detecting novelty. The question is, however, how do they do it? I discuss the question and suggest an answer to it. The story has its roots in previous work that I did some time ago with colleagues in Toronto, and also at the University of California at Davis, and that was significantly extended and elaborated by Lars-Göran Nilsson and his students in Stockholm. I first summarize this work and then raise and try to answer a question that emerged from it. It is now possible to imagine that the earlier version of the story may not have been quite right. Here, then, is an opportunity to make amends.
PETting memory

The story begins back in the 1990s, in the early heady days of “PETting memory”. The newly developed technique of positron emission tomography (PET) had just been adapted for studying human cognition and seemed to hold the promise of providing fascinating new insights into the mind and its relation to brain. At Toronto, too, we found ourselves learning how to play the new game with the new toy in the new sandbox. I and some colleagues constituted ourselves a “PETting memory” team and began to explore encoding and retrieval processes of episodic memory. In one of our early PET studies, we measured regional cerebral blood flow of subjects as they looked at complex coloured scenes of people and places (“travel pictures”) under two conditions: when the pictures were “new”, seen for the first time, and when they were “old”, having already been seen on the previous day (Tulving, Markowitsch, Kapur, Habib, & Houle, 1994b). When we subtracted the PET images for the new pictures from the images for the old pictures, we found a number of brain regions in the bilateral prefrontal cortex, left retrosplenial area, right angular gyrus, as well as some other cortical regions, that “lit up”, meaning that they were more active for the old than the new pictures. That was exciting, because it suggested that we were seeing the regions of the brain where “remembering happened” – a brand new facet of experimental memory research of which Ebbinghaus could not have even dreamt.

Our enthusiasm, however, was a bit dampened when we looked at the PET images made by the opposite subtraction – that is, when we subtracted the images for old (familiar, remembered) pictures from the images for new ones that the subjects had never seen before. Quite a few brain regions “lit up” in this contrast, too, especially when we decided to throw caution to the wind and used a low threshold for significance. This meant that when the owner of the brain remembered seeing the pictures that he had seen 24 hours earlier, some regions of the brain became less active. This was puzzling. What made the puzzle even more baffling was the fact that these “deactivations”, as they were called then (and sometimes are called so even today), occurred in what we thought could be properly classified as the “expanded limbic region”. This region included the hippocampus! (See Figure 6.1.) Thus, what we had observed was that remembering past experiences reduces the activity in the hippocampus, the “seat of memory”! This was nothing less than scandalous, although we could not say so publicly, of course. (As the informed reader can see, we were a bit more naive and innocent about the ways of memory and the brain than we are now. Among other things, we thought that all the changes in regional blood flow that one observes in a functional neuroimaging study necessarily reflect experimental manipulations!)

Early PETters of cognition did not like to see activity reductions (“deactivations”) in the brain because they did not quite know what to make of them. When task-related reductions in activity were observed, they tended to be just ignored or even not reported. By 1992, however, in an early PET study of
memory retrieval (Squire et al., 1992) that showed activation near the right hippocampus, a deactivated region in the right occipital cortex was also observed and was boldly interpreted as signalling “priming”, repetition-based facilitation of processing perceptual stimuli (Tulving & Schacter, 1990). Exactly how or why reduced neuronal activity should lead to improved behavioural performance was not quite spelt out – the question is still open, some 15 years later – but there was no problem with that: at the early stage of any scientific game, intuition often substitutes for logic. In our own very first PET study of what we thought was retrieval, this one with auditory sentences as the to-be-remembered material (Tulving, Kapur, Craik, Moscovitch, & Houle, 1994a), we had blithely copied the priming idea for the “deactivations” that had shown up, although we did not quite believe that all of the numerous “deactivations” signalled priming.

**Novelty in the hippocampus**

In the course of puzzling over the “unwanted” observation that the hippocampal formation tends to “slow down” when the subject looks at the remembered pictures, a happy, “novel” thought emerged that could be seen as “saving the phenomena” and fit reasonably into a world on which the hippocampus is central to memory. The thought was this: What if those
puzzling “deactivated” brain sites did not show decreased familiarity (remembering) activation for previously studied pictures but, instead, showed relatively increased novelty of the new pictures? Thus, what we had observed was not decreased familiarity but increased novelty. Logically, of course, the two propositions are indistinguishable, being but mirror images of each other. If the brain activity for perceived familiarity, in any given region, is lower than the activity for novelty, then it must also be the case that, in that region, the activity for perceived novelty must be higher than the activity for familiarity in that region. So, perhaps the hippocampus and its adjacent structures served memory, indeed, as advertised, but perhaps they did so by detecting novelty rather than by “remembering” what had happened. We found encouragement for this thought in relatively freshly published data on single-unit recordings from monkey brains that had revealed the existence of neurons that showed more spiking activity when the monkey looked at a picture for the first time, in comparison with another picture that was shown repeatedly (Fahy, Riches, & Brown, 1993; Li, Miller, & Desimone, 1993; Wilson & Rolls, 1990.)

So, we had PET data showing “familiarity activations” that presumably had something to do with the remembering of the pictures, and we had data showing “novelty activations” that presumably also pointed to a memory-related function. But what function? How does higher brain activity for novelty fit into the larger picture? It was well known at the time, of course, that novel events are better remembered than familiar events, but the connection of that fact with our finding of “novelty activations” in the limbic system that included the hippocampus was not at all obvious.

One thought that suggested itself was that distinguishing between what is novel and what is already known is useful to the brain when it comes to deciding whether or not to store incoming information for long-term use. If brains are clever – and as products of evolution they must be – they would not clutter up their memory stores with information that is already “in there”. Conversely, it is potentially advantageous to store information that is not yet “in there”, novel information, because it may turn out to be relevant in the future. So, clever brains distinguish between the old and the new in order to use their memory storage capacity economically. And the hippocampus would still exercise control over declarative memory (Squire, 1992), but would do so mostly by encoding through novelty detection, a critical component of encoding.

With some trepidation we published our new “discovery” of “widely distributed novelty-detection networks” (Tulving et al., 1994b) in a relatively new journal that was known for its reasonable reviewers and fast service to authors. Our message was that some brain networks are differentially involved in detecting novel stimuli, screening out the redundant information, and encoding the novel information in memory.

We specifically mentioned the hippocampus in the paper, although our “novelty activations” covered a much larger, albeit restricted, area than just
the hippocampus. In those early days of PETting the mind, the hippocampus seldom showed up on the otherwise fetchingly colourful brain maps. In a review paper, Richard Frackowiak, one of the leaders of the new field, praised the overall developments, because “localizing psychological models onto distinct parts of the brain provides a much sought-after link between metaphorical description and biology”, but he also pointed to the difficulties ahead, as “the hippocampus, long implicated in memory function, has been surprisingly difficult to activate” (Frackowiak, 1994, p. 111). So, lots of people were feverishly searching for the hippocampus in their PET studies of memory. Whenever a PET signal was detected anywhere near the medial temporal regions, it was warmly welcomed and eagerly labelled as “hippocampus”. We were part of the crowd when we reported novelty in the expanded limbic system in the right hemisphere, a huge area that included the hippocampus.

Right now this all sounds pretty weird, and certainly naive, but this realization is a sign of progress. And progress in PETting memory has been rapid. Today, with functional magnetic resonance imaging (fMRI) having largely replaced PET as the technique for neuroanatomical localization of mental activity, we know a great deal more than we did then, both about “memory in the brain” in general and about novelty/familiarity discrimination in particular. The hippocampus and its neocortical neighbours are now of equal interest to neurocognitive students of memory and are routinely observed in functional neuroimaging studies. “Back then”, the idea that the phenomenal feeling of oldness and newness of an event may arise from the activity of the limbic system, perhaps among other brain regions, was purely intuitive and therefore highly speculative. Today, however, we have comforting evidence of the connection (Aggleton & Brown, 2006; Daselaar, Fleck, & Cabeza, 2006; Gonsalves, Kahn, Curran, Norman, & Wagner, 2005; Nyberg, 2005).

**Novelty encoding hypothesis**

“Back then”, with the novelty detection networks revealed by PET (Tulving et al., 1994b), the question naturally arose as to the relation between the apparent ability of the limbic system to respond differentially to oldness/newness of stimuli and the well-known fact of superior memorability of novel stimuli. A possible answer to the question seemed to be that novel events are encoded more effectively, at the time of study, than familiar ones. The idea was formulated as the “novelty/encoding hypothesis” (Tulving & Kroll, 1995; Tulving, Markowitsch, Craik, Habib, & Houle, 1996). It postulated two sequentially organized sub-processes: novelty detection followed by “higher level” contextually determined encoding operations. The hippocampus and the limbic system would first serve to assess the novelty of the incoming information. Then, depending on the outcome of such assessment, the information would or would not be forwarded to other regions for the
higher level encoding. In another one of our early PET studies (Kapur et al., 1994) we had already caught a glimpse of some such “encoding regions”, including one in the left inferior prefrontal cortex. That one turned out to be a real winner, because today left inferior prefrontal cortex is acknowledged to be as important a player in the brain’s memory games as the hippocampus.

At any rate, the novelty encoding hypothesis holds that when incoming information is novel, it undergoes further encoding processing and ends up successfully stored. When it is totally familiar – for instance, one and the same picture presented for the umpteenth time in a long series – it is “screened out”, and its umpteenth presentation to the rememberer would not change what was already in the store before the umpteenth time. The two extremes, one of the highly novel and the other of the boringly familiar, are separated by a continuum of novelty detection and subsequent encoding, and a corresponding continuum of subsequent memory.

Now, if the novelty of an incoming item really is an important determinant of its encoding, one should be able to observe the effect even in purely behavioural tests. Furthermore, the effect should show up not just with pictorial material but with other kinds of incoming information as well. Because the behavioural data of our PET study were insufficient to test the idea – as we see below, we were missing an important control condition – I invited Neal Kroll at UC Davis to help me conduct an experiment whose design would allow a clean test of the novelty encoding hypothesis (Tulving & Kroll, 1995). We used what is known as the “3-phase recognition paradigm”, or simply “3-phase paradigm”. By holding constant all other variables that might affect memory performance, it allows complete control over oldness/freshness of stimuli.

Here is a sketch of the 3-phase paradigm:

<table>
<thead>
<tr>
<th>Phase</th>
<th>Study/Test</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1</td>
<td>Study</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 2</td>
<td>Study/Test</td>
<td>A</td>
<td></td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>Phase 3</td>
<td>Test</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

Each letter represents a set of individual items that subjects encounter. Thus, in Phase 1, subjects encounter items in experimental Sets A and B; in Phase 2, items in Sets A and C; and in Phase 3, items in Sets A, B, C, and D. All items “look alike” to the subjects; belongingness to different sets and subsets is known only to the experimenter.

The first two phases correspond to the classical yes/no recognition paradigm: in Phase 2, subjects are given a yes/no recognition test for items whose appearance they witnessed in Phase 1. In this test, items in Set A are the “old” targets, whereas items in Set C are the “new” lures.

In the 3-phase paradigm, Phase 1 becomes the “familiarization phase”. Its purpose is to create “familiar” items for the study list presented in Phase 2, in order to compare their learning in Phase 2 and subsequent retention with
“novel” (non-familiarized) items. Phase 3 provides the relevant data for this comparison. The subjects’ task is to discriminate between items studied (A and C) and those not studied (B and D) in Phase 2. The experimenter’s interest lies in whether the novelty/familiarity status of the test items affects such discrimination. The recognition test items in Phase 3 are A: familiar, studied; B: familiar, not studied; C: novel, studied; D: novel, not studied.

Thus, the design of the 3-phase paradigm is $2 \times 2$. One factor is familiarity/novelty, defined by the presence or absence of intra-experimental encounter prior to Phase 2. The second factor is study/non-study, defined by the item’s appearance or non-appearance in Phase 2.

The novelty encoding hypothesis “predicts” that subjects are more accurate at recognizing C (novel) items than A (familiar) items as having occurred in Phase 2. Accuracy of recognition can be measured as the difference between hit rates and false-alarm rates. (Using other measures, such as $d'$, does not materially change the outcome.) Thus the “prediction” of the novelty encoding hypothesis becomes one of interaction between the two factors of the $2 \times 2$ design. The difference between the probability of calling C items old and calling D items old should be larger than the difference between the probability of calling A items old and calling B items old – that is, $(C - D) > (A - B)$, where each letter represents the probability of items called “old” in the corresponding subsets (i.e., as having been seen in Phase 2).

Although we tested only four subjects, the experiment produced the predicted results with a vengeance. Recognition accuracy, as defined, was .20 for familiar words and .56 for novel words. This is a large difference. The $F$ ratio for the predicted interaction, with 1 and 3 degrees of freedom, turned out to be 280.3, which was “off the charts”, although we modestly reported it as having a $p$ value less than .01.

Physiologically oriented memory researchers took our “discovery” in stride, because it fitted snugly into the rapidly unfolding larger picture of brain and novelty. Psychologically oriented memory researchers were less keen on the study, and those few who seemed to have looked at the paper tended to be sceptical about both our findings and our thoughts about them (Chalmers & Humphreys, 1998; Dobbins, Kroll, Yonelinas, & Liu, 1998; Greene, 1999; Maddox & Estes, 1997). Among other things, doubts were expressed about the reliability and generalizability of the data. There were two earlier papers in the literature that reported the same kind of novelty effect that we had observed (Anderson & Bower, 1972, Expt. 4; Kinsbourne & George, 1974), but not too many people knew about them.

This is where Lars-Göran Nilsson led his troops into the breach (Åberg & Nilsson, 2001, 2003; Kormi-Nouri, Nilsson, & Ohta, 2005). In their experiments they adopted the general logic and design of the 3-phase paradigm and tested the effect of variables such as the format of the familiarization procedures, the nature of test materials, and the types of the activity between successive phases of the experiment. When they had finished, after several years of diligent work and spirited battles with critics, they had replicated the
How do brains detect novelty?

“novelty effect” many times. On a few occasions they had to make conces-
sions to the referees of their papers who had their own theories as to what was
going on in the 3-phase paradigm, and what the data really meant.

Thus, thanks to the effort of Lars-Göran Nilsson and his students, the
“novelty effect” as captured by the 3-phase recognition paradigm was firmly
established. However, a problem remained. The problem had been there
from the beginning, and its existence caused the researchers in Stockholm
considerable headache, because the critics of the work kept hammering away
at it: Well, you may have the data, and you can claim that they are caused by
novelty, but how do you explain them?

The critics were right: there was no explanation. The novelty/encoding
hypothesis offered no insight into why or how novelty detection, or novelty
assessment, occurs. It simply stated a generalization, derived from empirical
observations, that the brain (or the long-term memory system) responds
more vigorously to novel than to non-novel (recently encountered) stimuli
and that this enhanced vigour is translated into more effective encoding. It
was completely mute on any details of how this happens. This was not good.
To invoke the concept of novelty when reporting that people can tell the old
from the new, and that newness of stimuli enhances remembering, means
essentially to describe the 3-phase paradigm and the findings that it yields. It
does not say anything about how novelty detection “works”. To attribute the
effect of novelty to greater “attention” that subjects, at study, pay to novel
items also amounts to the description of the experiment and its results, with
an addition “epicycle” thrown in. It does not say how the subjects identify
novel items in order to pay more attention to them.

So, the problem remained: What are the underlying processes or mechan-
isms? To put it in slightly loftier language: what are the lathomena behind the
phenomena of novelty detection? This issue, explaining novelty detection,
will occupy us in the remainder of this chapter.

Comparator models of novelty detection

The most popular neurocognitive account of novelty detection at the present
time is provided by the “comparator model”, or the “match–mismatch”
model (Kumaran & Maguire, 2007b). Different versions of these models have
been proposed, but they share the central idea that novelty detection is based
on a comparison of the incoming information with information already
stored in the system. This comparison is carried out by specialized neuronal
networks, usually the hippocampus, or computational mechanisms, that
match incoming information against the information already in the system,
or against what the system (human, animal, robot), on the basis of its previ-
ous experience, “expects” to happen. If the difference between the actual and
the expected is sufficiently large there is a “mismatch”, and novelty is said to
have been detected. The statement in a recent paper by Kumaran and
Maguire (2007a) describing relevant fMRI data neatly captures these ideas:
“the hippocampus may generate predictions about how future events will unfold, and critically detect when these expectancies are violated, even when task demands do not require it” (p. 2372). The idea of the hippocampus as a novelty comparator has a respectable history that goes back to research and thinking before PET (Gray, 1982). Our PET findings of novelty detection seemed to fit into the larger picture nicely, even if our data showed novelty activations in a wider region, beyond the hippocampus. The notion of the hippocampus as a comparator became even more respectable when other PETters also reported sightings of the medial temporal regions when they compared PET responses to old and new pictures (Stern et al., 1996). An especially striking demonstration of the central role of the hippocampus was reported by Martin and coworkers when they compared old and new coloured, totally meaningless noise patterns (Martin, Wiggs, & Weisberg, 1997). Today the literature on the medial temporal lobe and novelty detection is extensive, and the critical regions that have been shown to be implicated have extended beyond the hippocampus into adjacent parahippocampal neocortical regions, especially perirhinal cortex (Brown & Aggleton, 2001), as well as the basal forebrain and the ventral tegmental area (Lisman & Grace, 2005).

Quite apart from the issue of “where in the brain” it happens, the idea that novelty detection requires a comparison between incoming information with that previously stored has been explicitly approved or implicitly accepted by many. In the Tulving and Kroll (1995) paper, we also took it for granted. Comparison seemed inevitably necessary. How else can anyone – the human rememberer, the brain, the hippocampus – decide whether a stimulus that one is facing now has not been encountered before, or at least not encountered recently, if not by comparing it with what has been encountered before? It is this apparently inescapable involvement of past experience in novelty detection that makes novelty detection a part of memory.

Can brains actually do it? Do they have to?

There is a problem. Hippocampus as a comparator is a metaphor. Explanations of novelty detection that are based on the concepts such as mismatch or comparison are no better, nor worse, than other metaphorical explanations that Richard Frackowiak, mentioned earlier, was talking about. Even when explanations are presented in the form of highly sophisticated models of how and where in the brain the relevant computations are carried out, they are still metaphorical. Kumaran and Maguire’s apt phrase that I quoted above about the hippocampus that generates predictions and detects violations of expectancies beautifully captures the spirit of this kind of explanation Kumaran & Maguire, 2007a, p. 2372). The description makes the central ideas clear, and as such it is heuristically most useful. But it need not be true.

Metaphorical explanations are among the time-honoured tools that brain/mind scientists use to fill gaps in their theories. They help us get on with the business. In many of these explanations, mind/brain processes or mechanisms
are likened to events and activities in our everyday world. The domain of memory is replete with anthropomorphic processes and mechanisms, beginning with various metaphorical renderings of “memory” itself (Roediger, 1980). When we talk about “strengthening” memory, we have in mind an analogy with strengthening of the muscle with practice; when we talk about retrieval as “search”, we think of the reading glasses or the car keys that we have misplaced in the house; when we talk about conscious awareness of our own experiences at times other than the present as “mental time travel”, and about the “self” that has the awareness as the “traveller”, we surely speak most metaphorically. In *Elements of Episodic Memory* (Tulving, 1983) I had decried the use of anthropomorphic descriptions of internal processes of memory because they created an illusion of understanding where we should have been admitting ignorance and using this ignorance as spur to further study. Among the examples of human-like activities pressed into service in encoding and retrieval that I listed were analysing, anticipating, categorizing, comparing, comprehending, deciding, discriminating, estimating, filtering, generating, interpreting, locating, marking, matching, mismatching, organizing, pigeon-holing, rehearsing, rejecting, scanning, searching, sorting, supplementing, switching, transferring, and understanding (Tulving, 1983, p. 141). In those days, the brain did not play much of a role in the study of human memory, and the processes were purely “cognitive”. Today, as the case of the hippocampus as the comparator shows, temptation is great to have the brain engage in these kinds of purely human activities too.

One problem with the comparator, or match–mismatch, model of novelty detection is that it leaves the fundamental problem unsolved: What is the mechanism that underlies novelty detection? To suggest that there is something like what we ordinarily mean by comparison actually going on in the brain means to stretch the credulity of many thinkers who distinguish between what the brain does, or can do, and what the whole organism, the brain’s owner, can do. It is difficult to imagine the physical brain actually comparing old and new objects. Oldness/newness is not a primary quality of objects – it is not like weight or length or hardness. It is not even a secondary quality – like hue or loudness or taste – which the brain, properly outfitted with relevant sensors, could be said to be able to detect and identify, in its own way, of course. Oldness/newness is a historical quality, and histories reside in the minds of thinking people, blessed with both highly evolved brains and educated, conscious minds.

The more important reason for questioning the idea of the brain’s ability to perform match–mismatch computations is that there is no need for such computations. It is not necessary to postulate any comparison mechanisms in order to account for novelty detection. There is no need to attribute to the physical brain (or the hippocampus, or its CA1 field) capabilities that they do not possess. Instead, there is a need to understand how the capabilities that they do have serve the brain’s owner.
I will argue that we can explain novelty detection in the brain without invoking comparisons and matches–mismatches. The brain can record happenings within the domains of its sensors and can store the information comprising these happenings. Various resulting records (engrams, memory traces) can become a part of the brain’s state at any given moment. This current state co-determines and modulates how any incoming information is processed, and it thereby shapes the consequences of that processing. In the case we are discussing here – novelty detection in the 3-phase paradigm – it is this current state that determines the differential processing of items in Sets A and C in the second phase of the 3-phase paradigm for their encoding and hence subsequent recognition. The point I am making is that it is possible to think of brain mechanisms that are involved in what the external observer of the brain calls novelty detection without attributing human-like capabilities to the human brain. One such possible mechanism is what I call “camatosis”.

The concept of camatosis

Camatosis is a (hypothetical) neurophysiological process that causes specific activity-dependent reduction in the efficacy of the operation of neuronal networks that support encoding of incoming information for long-term storage. The term is derived from “kamatos”, a word in classical Greek which can be translated into English as “tiredness” or “weariness”.

Camatosis – specific activity-dependent “weariness” of a neuronal network – is analogous to the “fatigue” that many biological systems incur as a consequence of their operations. Camatosis manifests itself in the diminution of the functional efficacy of a particular neural ensemble to do again at Time 2 what it did at an earlier time, Time 1. It applies to situations in which the presentation of an item (Event 1) occurs at time T1 and the presentation of the same, or similar, item (Event 2) occurs at a subsequent time T2. If the processing of Event 2 involves the same, or largely the same, neuronal networks that supported the processing of Event 1, and those networks are “camatotic”, processing of Event 2 will suffer – it will not occur “normally”, as it would in the absence of camatosis of the relevant networks. To the extent that this “processing” is responsible for encoding (conversion of a perceptual/cognitive event into a long-term memory trace), encoding of Event 2 will be impaired.

The concept of camatosis was initially introduced in a theoretical treatment of the inhibiting effects of earlier “learning” on subsequent “learning”, in situations where the interfering memory items were not (nominally) identical but only similar. In the very first publication on camatosis, the concept was used to account for the von-Restorff-type distinctiveness effects (Tulving & Rosenbaum, 2005). The idea was that what required explanation was not the good memory for the isolated item but, rather, the poor memory for the massed items. Camatosis was suggested as the “cause” of that poor memory. In a second paper, camatosis was proffered as an explanation, or as a factor
contributing to the explanation, of a variety of primacy effects in memory (Tulving, 2007). This paper also proposed a “law of camatotic encoding”: Of two events, the one whose encoding is more severely affected by camatosis is less likely to be retained. I will return to this “law” later.

The camatotic effect of Event 1 on Event 2 depends on many factors, as yet mostly unknown, although two relevant factors can be tentatively pinpointed on the basis of available evidence. One such is the similarity of Events 1 and 2: The greater this similarity, the greater the overlap between the neuronal networks processing Events 1 and 2, and the greater the camatotic effect of Event 1 on Event 2. Evidence from experiments conducted long ago and now nearly forgotten shows that in extreme cases of “massed repetition” (one and the same item presented for study twice in immediate succession), the additional opportunity for encoding afforded by the second presentation may have no effect at all (Madigan, 1969; Waugh, 1970). The other factor is the temporal interval between Events 1 and 2. Similar to the way the firing of a single spike in a single neuron is followed by a refractory period, the operation of a network is followed by a refractory period, although on a much wider timescale. A neat behavioural illustration of the refractoriness of what we now might refer to as camatosis can be found in Loess and Waugh (1967).

The camatosis hypothesis fits the 3-phase paradigm well: (nominally) identical items are repeatedly presented, and repetition impairs memory. In this paradigm, as described earlier, items familiarized in Phase 1 and presented again for study in Phase 2 (items of type A) are less likely to be identified as “old” in Phase 3, in comparison with novel items studied in Phase 2 (items of type C). What causes the difference? The answer is, camatosis. When a Set A item appears in Phase 1 of the 3-phase experiment, it is processed “normally” through the operation of a particular neuronal network whose “business” it is to register that kind of a particular input. When the same A item appears again in Phase 2, the event is again processed by very much the same network. This time, however, the processing cannot proceed quite “normally”, because some of the components of the network that were engaged in registering that item in Phase 1 have not yet recovered from their earlier operations. The less-than-optimal processing of the second occurrence of Item A produces a brain signal different from that of its first occurrence, and different from that of Item C when that item occurs, for the first time, in Phase 2. The observing scientist who compares these two signals notices the difference and says “Ha, here we have it, familiarity detection”, or “Ha, here we have it, novelty detection.”

The camatosis hypothesis

Even more relevant to our story in this chapter is the assumption that the impaired processing of Item A, when it is encountered by the subject in Phase 2, has a deleterious effect on the process of encoding that encounter, in long-term memory. It is this camatotically handicapped treatment of Set A items
that shows up in the $2 \times 2$ results of the 3-phase experiments (Åberg & Nilsson, 2001, 2003; Kormi-Nouri et al., 2005; Tulving & Kroll, 1995) as a difference between recognition accuracy of Set A items and Set C items. When we looked at these results “back then”, we could say “Ha, novelty effect in memory”, but, in the absence of an idea such as that about camatosis, we could not say anything much more about it. There was no explanation. Now we have a possible explanation. Novelty effects in memory, revealed by the 3-phase paradigm, represent a particular instance of the general “law of camatotic encoding”: Of two events, the one whose encoding is more severely affected by camatosis is less likely to be encoded and therefore less likely to be retained.

This kind of an explanation, derived from a “law”, may sound anachronistic in our day and age. Scientific laws were popular in psychology in the discipline’s early, formative years and decades (Teigen, 2002). Today they have all but vanished, for reasons that seem to be reasonably clear (Roediger, 2008). In this climate, to suggest a “law” seems almost heretical, especially if the temptation may be strong to glibly dismiss the proposed “law” as a definitional tautology.

The camatosis hypothesis could be illustrated graphically as shown in Figure 6.2. I have borrowed the illustration, with the authors’ kind permission, from Wiggs and Martin (1998). It shows how the network of features coding a perceptual (or any cognitive) object changes, or becomes “sharpened”, over its successive presentations. Wiggs and Martin created this neat diagram as an illustration of their own model of priming. Their caption was as follows:

![Figure 6.2 A possible schematic representation of the camatosis hypothesis. (Reproduced by permission from Wiggs & Martin, 1998.)](current Opinion in Neurobiology)
Illustration of the changes in a neuronal network representing visual object features as a function of repeated experience. As an object is presented repeatedly, neurons coding features that are not essential for recognizing the object decrease their responses (from black to grey to white), thereby weakening connections with other neurons in the ensemble (from black, to open, to no lines). As a result, the network becomes sparser and more selective, yielding enhanced object recognition. (Wiggs & Martin, 1998, p. 230)

For our present purposes – namely, to illustrate camatosis -- the caption might become something like this: “Illustration of the changes in a neuronal network representing features of a perceptual event as a function of its repeated occurrence. Some components involved in optimal (‘normal’) coding of the event become camatotic because of their previous engagement, as shown by changes from black, to open, to no lines, and the network becomes less efficient in representing the event.”

The camatosis model differs from the Wiggs and Martin model in that it assumes that the initial coding of the event is always optimal. In the Wiggs and Martin model it is not, insofar as it also involves non-essential features that are shed in the subsequent “sharpening” process. It also differs in that instead of the representation becoming “sharper” over successive repetitions, it becomes impoverished. These features of the camatosis model finesse the problem of why the initial coding of a stimulus is not as sharp as it could be, and why it includes nonessential features that can be dropped. Evolved brains have been practising object identification for hundreds of millions of years, and they ought to be pretty well tuned by now when it comes to the question of what to code and what not. At any rate, and regardless of whether we can ever know how clever brains really are, the camatosis model does not have any problem with dropped features of the network. All features that are needed for optimal (“normal”) processing are coded under “normal” conditions – that is, when the coding network is not camatotic by virtue of its recent engagement.

Note that according to the camatosis hypothesis there is no novelty detection in the brain, and there are no novelty detection circuits or networks. The brain does not classify the incoming information as novel, or familiar, or somewhere in between. And it does not screen out or forward to higher encoding networks information for further processing depending on its novelty/familiarity status. Our conjecture “back then” was wrong. Novelty detection exists, of course – otherwise we would not be talking about it – but it does not exist as a neural operation. It exists only as a concept, as an idea in the observing scientist’s mind. All that the brain does is to process old and new stimuli differently, in keeping with its current state, including the camatosis of the networks specific to different stimuli and different combinations of stimuli. It is the external observer who describes the products of such differential processing in comparative terms. The brain is a machine; it does
its thing (superbly, at least most of the time), but it cannot do things that the more powerful brain/mind system can.

The concept of camatosis is closely related to many other ideas that psychologists and other brain/mind scientists have created for the purpose of talking about what they do. There are the old perennials of habituation and adaptation that go back to the beginning of our science (Carandini, 2000; Sohal & Hasselmo, 2000). Numerous more specific terms have been used by researchers who have pursued problems in the neurophysiology of habituation and adaptation-like effects: “adaptive filtering” (Desimone, 1992), “adaptive mnemonic filtering” (Miller, Li, & Desimone, 1993), “stimulus specific adaptation” (Ringo, 1996), repetition suppression (Desimone, 1996; Henson & Rugg, 2003; Wiggs & Martin, 1998), decremental responses (Brown & Xiang, 1998), “repetition priming” or “neural priming” or just “priming” (Buckner & Koutstaal, 1998; Wagner, Koutstaal, Maril, Schacter, & Buckner, 2000), “cortical activity reduction” (Dobbins, Schnyer, Verfaellie, & Schacter, 2004), plus others. Recent developments point to the adoption of the term “repetition suppression” as a generic label of electrophysiologically and hemodynamically measured reductions of brain activity when a cognitively significant stimulus is presented repeatedly (Grill-Spector, Henson, & Martin, 2006).

Given this kind of embarrassment of riches, why yet another, new term? Why should anyone want to deliberately add to the terminological mess? There are at least two reasons. The first is relatively unimportant but worth mentioning nevertheless. Camatosis is the name of a hypothetical physiological mechanism that is introduced to explain, or help explain, particular behavioural/cognitive phenomena, such as novelty detection that leads to the novelty effect in memory. It is not a phenomenon itself; it is the hypothesized “cause” of a phenomenon. Making distinctions between causes and effects is healthy. Many terms that have cropped up even in my short essay – novelty detection, mismatch, adaptation, (repetition) priming – are ambiguous because they have been used in the sense of both observed behaviour and their underlying mechanism. It is an ancient problem in our science that persists, probably because scientists are busy people who have more important things on their minds than worrying about the cleanliness of the terms they use. Camatosis is supposed to be cleaner than many others in its circle of acquaintances, because it has a single meaning: specific activity-dependent “weariness” of a neuronal network subserving long-term memory encoding.

The second reason, a bit more substantial, has to do with camatosis not as a term but as a concept, an idea that has to do with things such as memory, repetition, and novelty detection. Like any other scientific concept, it receives its meaning and its usefulness by virtue of the role it plays in a wider theoretical context, and by virtue of its similarities to and differences from other concepts that figure in the same context (Dudai, Roediger, & Tulving, 2007). Camatosis is distinguished from other closely related concepts primarily by four ideas: (a) The insistence on the centrality of the fact that novelty is a
How do brains detect novelty? Strictly speaking, as I have suggested, they do not. Physical brains deal with the physical universe, and in the physical universe there is no such thing as novelty. This is why brains cannot “detect” novelty. Brains respond differentially to inputs, depending on their current state, which includes traces of past experiences. Intelligent observers can interpret some of the differences as reflecting novelty, but that is a propositionally expressed human judgment, not a brain decision.

In this chapter I have dealt with the problem of how to explain the observer-identified novelty detection as a necessary component of novelty effect in memory, the well-known phenomenon that novel events are remembered better than non-novel events. I limited the discussion to a special case of the novelty effect demonstrated in the 3-phase recognition paradigm, but it seems reasonable to expect that the same reasoning applies more generally to novelty effects of memory in other situations as well.

I noted that the popular comparator, or match-mismatch, models of novelty detection are unnecessarily complicated and attribute capacities to the brain as a physical/physiological “machine” that the brain probably does not possess. I also noted that “novelty” is a relational concept. To explain it means to explain the difference between the novel and the non-novel.

I described and discussed a recently introduced new concept, “camatosis” – the idea that the efficacy of neural networks involved in the process of encoding incoming information for long-term memory is reduced as a consequence of their own operations. Camatosis produced by earlier events deleteriously affects the encoding of subsequent events, depending on the extent to which the events in question are similar and are therefore subserved by correspondingly similar neural networks.

The concept of camatosis helps us understand the novelty effect observed in the 3-phase recognition paradigm – that items previously encountered in the experimental situation are less well recognized than items not previously encountered in the situation, or, in other words, that novel items are better remembered than familiar items. Camatosis leads to reduced efficacy of encoding of familiar items and thereby to their impaired retention in relation
to the non-familiar (novel) items. When an external observer, contemplating
the difference, focuses on the novel items, their higher memorability is seen as
the “novelty effect”. I suggested that this novelty effect can be viewed as a
particular instance of a general law of camatotic encoding: Of two events,
the one whose encoding is more severely affected by camatosis is less likely to
be retained.

The idea of camatosis as a co-determinant of encoding processes provides
only a broad and rough guide to exploring phenomena of the kind that led to
the novelty encoding hypothesis. Its practical merits are unknown and remain
to be determined by research. It claims the existence of physiological mechan-
isms that, automatically, inevitably come to restrain “normal” processes that
are involved in converting a perceptual object or event into a long-term
memory trace. The neurophysiological specifics of these mechanisms are not
specified and remain to be identified. This will happen in the future. The
important point is that even in its broad form the concept of camatosis allows
us to systematically relate phenomena of novelty and familiarity to other
interference phenomena of memory. When testing the camatosis hypothesis,
in real life or in laboratory experiments, however, it is imperative to mind the
ceteris paribus clause. It can be easily overlooked, yet it applies to the case of
camatosis as it does to all other cases of explanatory abstractions.

Acknowledgements

My research is supported by an endowment by Anne and Max Tanenbaum
in support of research in cognitive neuroscience.

Notes

1 The idea was not terribly original, as other thinkers had already discussed it
(Kohonen, Oja, & Lehtio, 1989, Metcalfe, 1993), but that did not matter, because
very few ideas ever arc. Although there was no experimental data directly relevant
to it, one thing that made it attractive was the fact that an obvious alternative
possibility that the brains store everything that happens and, every now and then,
somehow or other clears out the redundant or otherwise unneeded stuff – perhaps
at night, when the owner sleeps – sounded unnecessarily complicated, too much like
Rube Goldberg. An early selection mechanism seemed preferable, a bit more worthy
of the evolved brain.

2 In the context of Lars-Göran Nilsson’s Festschrift, an interesting detail of our
“travel pictures” PET study may be worth noting. The Talairach and Tournoux
(1988) coordinates of one of the “peaks” of novelty activation in the right medial
temporal region in that study were xyz = 26 −32 −8 (Tulving et al., 1994b, Table 1).
In an fMRI study, co-authored by the celebree of this Festschrift and its two editors
(Lind et al., 2006), the same hippocampal region (xyz = 30 −30 −8, MNI [Montreal
Neurological Institute] coordinates) was identified as one in which a similar novelty
activation (novel > familiar) for words was observed for one of two groups of
APOE carriers (ε3/3), while the other group (ε4) showed activation in the reverse
direction (familiar > novel). APOE ε4 carriers are known to be at high risk for
Alzheimer’s disease and impairment of episodic memory, hence the potential
significance of the observation. But regardless of that implication, the close match between the “novelty activations” in the two studies, widely separated geographically and historically, is “cool”.

3 As noted earlier, one can write things in a Festschrift that diligent censors would cross out elsewhere. “Lathomena” (plural of “lathomenon”) is a real word that one can find in a real book (Tulving, 1983, pp. 123, 349), as well as in Simple English Wikipedia on the Internet, if you look up “phenomenon” there, although it is not in any dictionary, yet.

4 I am grateful to Jaan Puhvel for creating the term for its intended purpose.

5 As brains are permanently active, in one way or another, they are also permanently camatotic, in one way or another. What I refer to as “normal” processing of given input refers to processing under the conditions where there has been no recent brain activity involving components of the specific functionally relevant networks underlying the processing of that particular input.

References


Maddox, W. T., & Estes, W. K. (1997). Direct and indirect stimulus-frequency effects


Novelty and familiarity activations in PET studies of memory encoding and retrieval. *Cerebral Cortex, 6*, 71–79.


