Effective Categorization of Objects, Scenes, and Faces Through Time

Éric McCabe, Caroline Blais & Frédéric Gosselin.
Département de psychologie, Université de Montréal

Key words: categorization; attention; vision; temporal processing; object, face and scene recognition.

Abstract

We first describe *SLIP* (Strategy Length & Internal Practicability) a formal model for thinking about categorization, in particular about categorization through time. We then discuss an early application of this model to basic-levelness. We continue with novel evidence for discrete processing cycles through time. We then turn to aspects of categorization through time that have been neglected in the categorization literature: our limited processing capacities; the necessity of having a flexible categorization apparatus; and the paradox that this inexorably brings about. Finally, we spend several pages on a two-fold resolution of this paradox. Throughout, but especially toward the end, we attempt to bridge work done in categorization, vision, neuropsychology, and physiology.

1. A model of categorization

Figure 1 shows four artificial scenes synthesized by combining two different luminance patterns (that we call *flat* and *hilly*) with two different chromatic patterns (labeled *grassy* and *sandy*). This toy-example captures some of the essential characteristics of real-world categorization. These stimuli can be categorized as either "field" (the combination of *is_flat* and *is_grassy*), "desert" (*is_flat* and *is_sandy*), "mountain" (*is_hilly* and *is_grassy*), or "dune" (*is_hilly* and *is_sandy*) at the most specific level of categorization. At the most abstract level of categorization there are more than one possibilities – we will come back to this point latter – but, for now, let us only consider the two categories "flat" and "hilly" (Figure 1). Thus, the "mountain" and the "dune" scenes are also "hilly", and the "field" and "desert" scenes are also "flat". We thus have a small category hierarchy. And we are ready to begin the unpacking of our ideas about categorization through time.

Insert Figure 1 about here

SLIP (Strategy Length & Internal Practicability) is a categorizer that applies "optimal" testing strategies to determine the category membership of objects. The goal here is not so much to mimic human performance precisely but to provide a non-arbitrary starting point for future modeling efforts as well as a framework to better understand human performance (e.g., Anderson, 1990, 1991; Kersten, 1990; Feldman, 2000). A strategy comprises sets of noisy detectors. For example, the desert scene illustrated in Figure 1 satisfies two category strategies: Strat("dune") = [{*is_hilly*} & {*is_sandy*}], is the SLIP strategy for the "dune" category and comprises two sets of detectors; Strat("hilly") = [{*is_hilly*}] is the SLIP strategy for the "hilly" category and comprises a single set of detectors. We think of these sets of detectors as populations of specialized neurons (e.g., in V4 for color, in V5 for motion). SLIP launches *a subset* of all these detectors in parallel. The size of this subset is related to the amount of information that humans can process simultaneously. We discuss this point in some details in a subsequent section.

Because the detectors in a set are redundant, only one of them needs to be successful to verify the entire set. For example, to verify that a scene is "flat" in the category hierarchy illustrated in Figure 1 one successful luminance detector suffices. Everything else being equal, SLIP predicts that strategies associated with more redundant sets of detectors will have a higher probability of being completed after few discrete processing cycles (*t*). There are two ways of increasing the redundancy of a strategy: either more detectors of feature *X* become available, or more exemplars of feature *X* become available.

Often more than one set of redundant detectors is required to place a scene in a category. For example, to verify that a scene is a “dune” in the category hierarchy displayed in Figure 1, one successful luminance detector and one successful chrominance detector are required. Everything else being equal, SLIP predicts that strategies associated with shorter strategies will have a higher probability of being completed after few processing cycles.

We now turn to the formalization of these ideas. The cumulative probability that a strategy comprising *n* sets of redundant detectors is completed at processing cycle *t* or before is given by:

$$= \overset{n}{\underset{j=1}{}} \left(1 \quad {}^{t}_{j}\right) \qquad \text{(Equation 1),}$$

with ${}_{j} = 1 \quad {}_{j}$. The constant ${}_{j}$ is the probability that the set of detectors *j* is successful after one processing cycle. It is the “weight” given to dimension *j* (for details see Gosselin & Schyns, 2001b).

Although the probability distribution of Equation 1 is useful in its own right, we have more often employed its associated density function. That is, the function that gives the probability that a *SLIP* strategy is completed after *exactly t* processing cycles. To compute it, we must subtract two cumulative probabilities: the probability that the a strategy is completed in at most *t* processing cycles minus the probability that it is completed in at most *t-1* processing cycles:

$$= \quad (t) \quad (t \quad 1) \qquad \text{(Equation 2).}$$

To illustrate Equation 2, we have applied it to Strat(“hilly”) and to Strat(“dune”). The predicted density functions are given in Figure 1. Here, we have assumed that both ’s (i.e., the probability that the set of detectors *j* is unsuccessful during one processing cycle) were equal to .5. Two things are remarkable about these density functions: (1) They are very different from one another. In particular, they differ on the average number of cycles necessary before verification. The means of such density functions is equal to $\overset{+}{\underset{t=1}{}} t \quad (t)$, i.e. 2 cycles for Strat(“hilly”) and 2.67 cycles for Strat(“dune”). (2) Another remarkable aspect of these density functions is their shapes which are reminiscent of response time density functions (e.g., Luce, 1986). If we assume that response time (*RT*) is a linear function of the number of processing cycles (i.e., *RT* = *a* * *t* + *b*, with *a* and *b*, two free parameters), Equation

2 can be construed as an RT density function. Applying this assumption to the above example, this implies that categorizers will take longer to verify that a scene is a “dune” than to verify that it is “hilly”. This connection between average RT and average number of processing cycles gives us a first quantitative handle on the experimental literature in cognitive psychology. We will see in the next section how SLIP can explain a large chunk of the so-called basic-level literature.

2. The basic-level literature

In Rosch, Mervis, Gray, Johnson and Boyes-Braem (1976, Experiment 7), participants were taught the names of 18 objects at three levels of categorization–the *subordinate* (e.g., “Levis”, “Macintosh”), *basic* (e.g., “pants”, “apple”) and *superordinate* (e.g., “clothes”, “fruit”). These objects belonged to one of six possible non-animal taxonomies: “musical instruments”, “fruits”, “tools”, “clothes”, “vehicles”, and “furnitures”. In a verification task, subjects were shown a category name followed by a stimulus picture, and had to determine whether they matched. On average, categories at the basic-level were fastest to verify, and categories at the subordinate level slowest (see also Hoffmann & Ziessler, 1983; Jolicoeur, Gluck & Kosslyn, 1984; Murphy, 1991; Murphy & Smith, 1982; Murphy & Brownell, 1985; Tanaka & Taylor, 1991). This is the first of many verification experiments that demonstrated a superiority at basic level of abstraction. The SLIP model is perfectly adapted to predicting *basic-levelness* that is, the average speed of categorization at various levels of abstraction in a verification task.

Gosselin and Schyns (2001b) compared the predictive power of SLIP with that of four other basic-level measures: context model (Medin & Schaffer, 1978; modified by Estes, 1994), category feature-possession (Jones, 1983), category utility (Corter & Gluck, 1992), and compression measure (Pothos & Chater, 1998a, 1998b), drawing data from the empirical work of Rosch et al. (1976), Murphy and Smith (1982), Mervis and Crisafi (1982), Hoffmann and Ziessler (1983), Corter, Gluck and Bower (1988), Murphy (1991), Lassaline (1990), Tanaka and Taylor (1991), Johnson and Mervis (1997), Gosselin and Schyns (1998), and three novel experiments using computer-synthesized 3-D artificial objects. SLIP led the pack by a large margin predicting 88% of this (ordinal) data set, category utility 64%, category feature-possession 62%, the compression measure 42%, and the context model 35%.

A critical aspect of basic-levelness is that it optimizes a number of indexes of performance. Convergence of all of these is crucial to establish a preferred categorization level, even though verification speed is the most commonly used. It was thus important for

Gosselin and Schyns to show that SLIP is not limited to model category verification and they did; however, this goes beyond the focus of this chapter. The interested reader is referred to Gosselin and Schyns (2001b), where most of what we have so far written was originally presented. Next, we will give a new twist on this relatively old story. We will look at the way categorization occurs through time at the atomic level. This is an aspect of the question that has been completely neglected in the categorization literature.

3. Discrete processing cycles

SLIP proposes that we apprehend the world via discrete processing cycles (i.e., the *t*'s discussed above). This is in clear disagreement with the apparent continuous nature of our every-day experience of time. We and others, however, have recently gathered compelling evidence that our visual world does in fact tic (e.g., Van Rullen & Koch, 2003b; Ward, 2003). Perhaps the most direct of these evidences comes from a recent study carried out in our laboratory using a powerful experimental technique called *Bubbles* (Gosselin, Lapalme & McCabe, in preparation). We will describe this experiment in a somewhat detailed manner in the following section. This section can also be understood as a *Bubbles* primer; it will also be useful because the technique is at the heart of several other experiments reported in this chapter.

3.1 A Bubbles primer

Bubbles is a generic procedure that can reveal the information that drives a measurable response (Gosselin & Schyns, 2001a). Six decisions are required in order to set up a *Bubbles* experiment: 1) what is the stimulus set, 2) in which space will stimuli be generated, 3) what is the "bubble," 4) what is the observer's task, 5) what are the observer's possible responses, and 6) is the analysis per observer, or per group of observers (Gosselin & Schyns, in press). Next, we discuss each of these decisions in the context of Gosselin et al. (in preparation).

3.1.1 Stimulus set? In a *Bubbles* experiment, the stimulus set is crucial because it critically bounds what will be tested. Gosselin et al. used 640 natural scenes equally divided into eight categories (i.e., "city", "highway", "road", "mountain", "field", "beach", "houses", "forest"). The scenes were gray shaded images of resolution 128 x 128 pixels (subtending 2.8 x 2.8 deg of visual angle). The overall energy of the scenes was normalized. These scenes were presented for a duration of 150 ms. Generally speaking, the larger the stimuli set, the better the *Bubbles* solution should be. A large stimulus set will tend to prevent observers from adopting strategies atypical of natural processing.

3.1.2 Stimulus Generation Space? The choice of a proper stimulus generation space is one of the most important decisions when setting up a *Bubbles* experiment. Gosselin et al. searched the temporal dimension with a resolution of 120 Hz. We will briefly describe other stimulus generation spaces in section *5.1.1*.

3.1.3 The samples? At this stage, two important decisions have been made and the search can almost begin. In the search, information is sampled from the set up space, and the next decision to make concerns the unit of sampling. This unit depends on a number of factors, including the stimuli, the nature of the search space and the task to be performed. Here we varied the contrast of the scene through time. The contrast of the scenes presented by Gosselin et al. was modulated by a vector of Gaussian white noise low-passed with a Butterworth filter at 30 Hz (to prevent flicker fusion). Accuracy was maintained at 50% correct by adjusting the maximum contrast of the scenes using a gradient descent algorithm.

3.1.4 The task? At this stage, the sampling procedure has been fully specified. Another important decision is that of the task. In Gosselin et al., it was a categorization task. Participants had to put the various scenes into the proper categories. Several tasks can and have been used with the same stimulus set.

3.1.5 Response? The response is an interesting parameter of a *Bubbles* experiment because the technique is in principle sensitive to any measurable dependent variable. In Gosselin et al., observers pressed labeled keys corresponding to the eight scene categories. Such key-press responses have been used to derive correct and incorrect responses (e.g., Bonnar, Gosselin & Schyns, 2002; Gibson, Wasserman, Gosselin & Schyns, in press; Gosselin & Schyns, 2001a; Schyns, Bonnar & Gosselin, 2002; Vinette & Gosselin, 2002), and response latencies (Schyns et al., 2002). Electroencephalographic (EEG) activity have also been used in past experiments (e.g., Schyns, Jentzsch, Johnson, Schweinberger, & Gosselin, 2003; Smith, Gosselin & Schyns, 2004). Other responses could be the firing rate of single cells, fMRI, galvanic skin response, plethysmograph, eye movements, and so forth. To the extent that *Bubbles* is essentially an empirical tool, it is useful to record as many different responses as possible (e.g., correct/incorrect, latencies and EEG in a face recognition experiment). It is difficult to predict before the experiment how responses will correlate with the parameters of the search space..

3.1.6 Observers? Depending on the objectives of the research, different types of observers can interact with the bubbled stimuli. Gosselin et al. used human observers. Each performed about 3,600 trials. Brain lesion patients and animals have also performed *Bubbles* experiments as will be seen in section 5.1.1

3.2 Fossilized discrete processing cycles

Once the search has been run and the data are collected, the analyses can be performed. The goal of the search is to isolate a subspace of information correlated with the measured response(s). Typically, a multiple linear regression on the samples (explanatory variable) and the responses (predictive variable) provides this solution. For the study of Gosselin et al. (in preparation) this analysis reduced to summing all the filtered noise vectors that led to a correct response in a first vector – the correct vector – and all filtered noise vectors, irrespective of accuracy, in a different vector – the total vector – and to dividing the correct vector by the total vector element per element. The result is a vector giving us the probability that if a temporal slot is sampled a correct answer will be reached in this experimental setting, i.e. a *time classification image*. The first column in Figure 2 illustrates the outcome of this analysis on ten subjects.

Insert Figure 2 about here

The most striking feature of these plots is the presence of a pulsation in the range of 20 to 27 Hz (i.e., within the beta bandwidth). We believe that these are fossilized atomic discrete processing cycles. Another pulsation is present in some classification images (this is especially obvious in the time classification images of subjects S2, S5 and S10). This additional pulsation, however, has a period just about an integer multiple (either 2 or 3, depending on the observer) of that of the fastest pulsation; to us, this suggests that on some trials – and for some yet unknown reasons – subjects “missed” one processing cycle out of two or three.

In a related vein, Vinette et al. (2004) used a space-time version of *Bubbles* in order to examine how, during a face identification task, visual information is extracted from stimuli in the first 280 ms after their onset. They obtained a clear pattern of results: The eye on the left side of the image became diagnostic between 47 and 94 ms after the onset of the stimulus; after 94 ms, both eyes were used effectively. More relevantly, Vinette et al. were the first to observe a sinusoid pulsation in the range of 7 to 14 Hz (i.e., within the alpha bandwidth) in the effective use of temporal information. They could not reveal beta pulsations because they did not have the temporal resolution required to do so.

In order for time-locked classification images to reveal anything at all, time slots need to possess special values. This requirement was satisfied in the two experiments just described because the processing cycles they revealed are 1) regular and 2) in phase with stimulus onset[1]. To a SLIP categorizer, however, temporal slots do not have any special meaning; and thus we cannot learn anything about a SLIP categorizer with time classification images. In the next section, we will show how the data from a *Bubbles* experiment can be analyzed differently and inform us about a SLIP categorizer.

3.3 What can temporal bubbles reveal about a SLIP categorizer?

So far the *Bubbles* procedure has only been modeled for the Linear Amplifier Model (LAM) observer in the spatial domain (Murray & Gold, 2004). Here, we sketch an answer to the question: What can temporal bubbles reveal about a SLIP categorizer?

We will assume that a SLIP detector can fail at time t either because it is noisy or because no bubble reveals information. In Gosselin et al.'s (in preparation) experiment, a sample either reveals or does not reveal information at a particular time slot with equal probability. The behavior of a SLIP categorizer during such an experiment can be described by replacing the $_j$ in equations 1 and 2 by $_j = .5\ _j$. Apart for this "slowing down", we will suppose that the categorization process of a SLIP categorizer is unaltered by the sampling occurring during a *Bubbles* experiment (for a thorough discussion of this issue see Gosselin & Schyns, 2004).

What really matters to a SLIP categorizer then is the number of processing cycles during which information was revealed. Suppose, for example, that on a particular trial, one bubble falls on the first processing cycle and another bubble falls on the fourth processing cycle. Information is thus revealed for a total of two processing cycles. During other trials, information will be revealed for a total of 1, 2, 3, … processing cycles. And, for each of these information slots, it is possible to compute the probability of a correct response. We will call the resulting vector of proportion correct an *information classification image.* We mentioned at the beginning of this chapter that the cumulative probability that a SLIP strategy comprising n sets of redundant detectors is completed at processing cycle t or before – i.e., the expected information classification image of any given SLIP categorizer – is given by Equation 1. The meaning of t in the Equation must however be slightly modified: "tth processing cycle" must be replaced by "tth revealed processing cycle".

---

[1] *SLIP* postulates processing "spikes" rather than the "sinusoid" oscillations observed. However, given some phase uncertainty (e.g., modeled by the convolution with a Gaussian function), the former can be made to mimic the latter.

In the *Bubbles* experiment of Gosselin et al. (in preparation), information classification images can be estimated as follows: The number of processing cycles available to an observer on trial *t* is proportional to the sum of the elements of the filtered Gaussian noise vector employed to sample visual information on that particular trial weighted element by element by that observer's time classification image (see previous section *3.2 Fossilized discrete processing cycles*). The second column in Figure 2 displays the information classification images extracted using this procedure on the ten subjects of Gosselin et al.. In theory, once we have extracted the empirical information classification image of a human categorizer, we can bestfit it to Equation 1 (with the $_j$, not the $_j$) and the parameters that minimize the error are estimates of the SLIP categorizer closest to that particular human being. Before we can apply this scheme practically, however, a tremendous quantity of work will need to be done. For example, we will have to study just how *identifiable* a SLIP categorizer is given an information classification image. To be continued…

4. The need for flexibility and a paradox

So far our story has been to a large extent at least a success story. We will now turn to limitations of the SLIP model and models like it and to ways to overcome these limitations. We hope that this will point toward new directions for research in categorization unfolding through time.

4.1 Limited processing capacity

There is a long and venerated tradition of research on the topic of information processing capacities in the field of human cognition (e.g., Broadbent, 1958). We will not dwell much on this vast literature here. It will be enough for our purpose to cite a few representative examples. Most of the experiments in this field demonstrate one way or another that human information processing capacities are far less impressive than what humans would naïvely expect. In a seminal article, Miller (1956) showed that our short term memory has a capacity of seven give or take two "chunks" of information. Similarly, "object tracking" experiments performed by Pylyshyn and colleagues have shown that we can only track four or five moving targets simultaneously (e.g., Sears & Pylyshyn, 2000). The most striking demonstrations ever perhaps come from so-called "change blindness" experiments. Observers are asked to detect important changes in a natural or an artificial scene and are shown to be ridiculously poor at it (e.g., Rensink, O'Regan & Clark, 1997; Simons & Levin, 1997; Simons, 2000b (see the special issue of Visual Cognition); Most, Scholl, Clifford & Simons, in press). In the related paradigm of "inattentional blindness" (Mack & Rock, 1998; Simons, 2000b), observers are asked to perform a task that, unbeknownst to them, is a distraction task (e.g., to count the number of times the members of team in white T-shirts pass a basketball). In small portion of the trials something different happens (e.g., a human dressed in a gorilla suit walks to the center of the scene, turns toward the spectators, beats its chest with its hands a few times, and walks away). Usually less than one fourth of the participants notice these odd trials (Mack & Rock, 1998; Simons, 2000a).

As we have written *en passant* the SLIP model implements this psychological reality by having only a subset of all available detectors activated simultaneously. So if we are blind to a large change occurring in a natural scene – say an engine disappearing from the wing of an airplane – it is because not enough of the relevant detectors are active to complete the verification on time; and it suffices to activate more of these relevant detectors to see the change.

4.2 The need for flexibility

The limited processing capacity of the cognitive system implies that a selection of information must occur. There is now a wealth of evidence that this does in fact happen. We will review some of the most compelling empirical evidences for this in section *5.2 Flexible iterative processing sweeps*. As a preview, here we will consider an experiment performed with the four artificial scenes of Figure 1 (Gosselin & Schyns, submitted). In a learning phase, all participants learned to categorize the four scenes at a general and at a specific level. At a general level, the LUMI participants learned to separate the four scenes into "flat" and "hilly" on the basis of luminance cues; and the CHRO participants learned to separate the same scenes into "grassy" and "sandy" on the basis of chromatic cues. At a specific level of categorization, both LUMI and CHRO participants learned to categorize the stimuli as either "field" (the combination of *is_flat* and *is_grassy*), "desert" (*is_flat* and *is_sandy*), "mountain" (*is_hilly* and *is_grassy*) or "dune" (*is_hilly* and *is_sandy*). In a testing phase, participants were instructed to categorize the scenes at their most specific level (never at their general level). Note that the specific categorizations are strictly identical in the groups, which only differ on the dimension structuring their high-level categorizations. The conjunctive nature of the stimuli can be used to determine *indirect* effects of diagnosticity. In the context of the SLIP categorizer, Gosselin and Schyns predicted that CHRO observers would weight the chrominance dimension more heavily than the luminance dimension, whereas LUMI observers weighted the luminance dimension more heavily than the chrominance dimension[2]. This would happen if each group was tuned to chromatic and luminance information to maximize their categorization potential. After a successful test only on the luminance (vs. chrominance) dimension, the LUMI (vs. CHRO) group can already categorize the scene at a general level whereas the CHRO (vs. LUMI) group cannot.

Here this does not help the participants much but in real life putting an object in a category allows them to infer unseen features (e.g., Rosch, 1978; Anderson, 1990). Consider the example of Pi Patel, the main character in *Life of Pi*, on his raft with an unknown thing. At least two categorization routes of similar processing time can lead him to the same conclusion: "Richard Parker" is standing just in front of me. The first route would initially verify that the thing possessed *distinctive tiger marks* (property *a*), then the *specific eye color of Richard Parker* (property *b*). The second route would perform the same property tests, but in the opposite order (*eye color* before *distinctive tiger marks*). Both routes lead to the same final outcome in the same time: "This is Richard Parker, a tiger". However, in the first route, the initial testing of the *tiger marks* provides an intermediate "tiger" categorization before the specific "Richard Parker", allowing Pi Patel to react faster. This intermediate categorization arises from the generic knowledge that tigers have *tiger marks*. However, in this struggle, perspective matters. "Richard Parker" would probably prefer the second route, for an effortless dinner. In contrast, evolutionary arguments would probably select the first route: a "tiger" is after all a *man-killer.*

Very little is known about how categorizations are embedded in real life. Gosselin and Schyns' experiment suggests however that beyond the probabilistic preference for categories with high basic-levelness induced by their feature structure, we are biased for a sequence of categories. There is actually some indication of this in the face recognition literature: Liu, Harris and Kanwisher (2002), after having conducted a MEG study, proposed that face perception should be divided into stages: a first stage where the stimulus is categorized like a face (occurring in the first 100ms) and a second stage, completed after about 170 ms, where the face is identified at an individual level. Sugase, Yamane, Ueno and Kawano (1999) have obtained similar results performing unicellular recordings in macaque.

---

[2] Gosselin and Schyns (submitted) adapted SLIP to predicting the error patterns of subjects in this situation. In a nutshell, they used Equation 1 and corrected it for guessing. The average bestfits for the various types of response are reproduced in Figure 1b. Importantly, observer groups assigned orthogonal weights to the luminance and chrominance dimensions (with the CHRO vs. LUMI group biased to the chromatic vs. luminance dimension, with greater weights of about .6) even though categorizations at the specific level (the task to resolve) was itself unbiased to one or the other dimension.

These authors showed that information contained in a face is used in a first rapid stage where global information allow the discrimination between macaque and human face and a second slower stage where finer information is used in order to identify face or categorize facial expression.

4.3 Back to the paradox

We have seen how the limited processing capacities of humans necessitate flexibility, and that this flexibility has, to some degree, now been incorporated into categorization models. However, the above SLIP story and ones like it, for all their appeal, inexorably lead to a paradox: How can a categorizer know which detectors to turn on before knowing what is out there? And, reciprocally, how can a categorizer know what is out there before turning the relevant detectors on? We will devote the next few pages on ways to resolve this paradoxical situation.

5. Categorization as an iterative process

Our resolution of the above paradox will be two-fold: We will argue that a subset of a SLIP categorizer's – or any other categorizer's – detectors is *always* activated and that the remainder is used in a flexible manner, informed by previously activated detectors. Many theoretical proposals in the visual recognition literature are in line with this answer. We will review these proposals, with categorization always on our minds. We will also describe a portion of the empirical work that supports these theories.

5.1 Compulsory feedforward processing sweeps

Ullman (1984) was among the first to propose that object recognition is informed by feedforward compulsory processing sweeps (or *visual routines* as he called this theoretical construct). Some detectors – using the terminology of the SLIP framework – would always be activated and would thus allow for surprise, for unexpected things to be discovered. Of course, these detectors cannot fully categorize the visual scene; otherwise, we would be back to square one paradoxically speaking. This compulsory feedforward processing sweep, however, can attract our attention – guide the activation of our flexible detectors – toward suspicious-looking, partially processed objects.

Numerous studies performed by Thorpe and his research group as well as by others demonstrate the capacity of human subjects to categorize visual scene very rapidly. Assuming that flexibility and feedback require time, this rapid processing supports the claim that a compulsory feedforward sweep can perform relatively complex processings (e.g., Delorme, Richard & Fabre-Thorpe, 2000; Fabre-Thorpe, Delorme, Marlot & Thorpe, 2001; Thorpe, Fize & Marlot, 1996; VanRullen & Thorpe, 2001a). It has been demonstrated, for instance, that human participants can categorize a natural scene flashed for 20 ms with high accuracy (94%; Thorpe et al., 1996). In such *ultra-rapid categorizations*, a differential electrophysiological component distinguishes target from non-target scenes around 150 ms following stimulus onset (Thorpe et al., 1996); this brain activity is also correlated with the subject's decision about the status of the stimulus (i.e. target vs non-target; VanRullen & Thorpe, 2001c). It seems likely that the mechanisms involved in ultra-rapid categorization are purely feedforward and encapsulated (Fabre-Thorpe et al., 2001; VanRullen, Delorme & Thorpe, 2001; VanRullen & Koch, 2003a): The performance of subjects in such tasks does not increase with training even for period as extensive as 14 days (Fabre-Thorpe et al., 2001). Furthermore, a purely feedforward biologically-inspired neural network was shown to be sufficient to duplicate ultra-rapid categorization performance in humans (Delorme & Thorpe, 2001; VanRullen, Gautrais, Delorme & Thorpe, 1998).

Likewise, it was shown that ultra-rapid categorization of natural scenes (Li, VanRullen, Koch & Perona, 2002) and gender discrimination of faces (Reddy, Wilken & Koch, 2004) are possible in the near absence of attention. The performance of human subjects in both of these tasks was shown to be unimpaired by a dual-task requiring attention. Using a battery of experimental tools, LeDoux and colleagues (e.g., Armony & Ledoux, 2000) have shown that fearful faces are processed in a fast and feedfoward manner by the amygdala and can subsequently drive behaviour. One question that remains unanswered by all these experiments concerns the nature of the information processed during these compulsory feedforward sweeps.

5.1.1 The nature of the information processed during compulsory feedforward sweeps

Adolphs, Gosselin, Buchanan, Tranel, Schyns and Damasio (2005) used the *Bubbles* technique (see section *3.1 A Bubbles primer*) to pinpoint the nature of the information processed by the amygdala during a fearful versus happy face discrimination task. They compared the information used effectively to recognize emotion through facial expression by SM, a rare female patient with complete, bilateral damage restricted to the amygdala, with the information used effectively by 10 matched controls. Each subject was submitted to about 3,000 trials. In a trial, one of four faces was transformed in an image generation space comprising three dimensions (i.e., the standard X and Y axes of the image plane, plus a third

Z axis representing spatial frequencies), and was sparsely revealed by randomly located Gaussian holes. This sampling procedure is illustrated in Figure 3a. Adolphs et al. then performed multiple linear regression on the location of the Gaussian holes and response accuracy to determine which areas of the image generation were used most effectively. These regression coefficients were Z-transformed, thresholded at 1.65 ($p < .05$), and used to construct effective faces. The results are illustrated in Figure 3b. They clearly show that the amygdala is involved in the processing of high spatial frequency eye information.

Insert Figure 3 about here

In another study using the same *Bubbles* search space, Bacon, Vinette, Gosselin and Faubert (2003) examined which type of information maximizes conscious and unconscious priming (they used the task devised by Dehaene et al., 2001). Their results indicate that conscious priming depends mostly on high spatial frequency, local features whereas unconscious priming depends mostly on global features in the lower spatial frequency range (see Figure 3c). This, as you will learn in the next section of this chapter, is perfectly consistent with the recent theoretical proposals of Bar (2003) and Bullier (2001a; 2001b).

5.2 Flexible iterative processing sweeps

The previous section reviewed the evidence for the existence of compulsory feedforward processing sweeps or, transposed into the SLIP framework outlined above, for the compulsory activation of some detectors. Compulsory feedforward sweeps fit marvelously well within the "standard" feedforward anatomical hierarchy of the visual system (Felleman & Van Essen, 1991). It is easy to understand why this has led mostly to bottom-up and constructivist models of information processing in the brain (Biederman, 1987; Marr, 1982). Recent experiments, however, have seriously challenged this viewpoint by showing the crucial importance of top-down processing (see also section *4.2 The need for flexibility*). These studies have thus set the stage for new explanatory models comprising either simple top-down components, or complicated iterative loops. We will present physiological evidences for the existence of iterative loops as well as three models that mix early compulsory feedfoward processing sweeps with late iterative processing sweeps.

According to Bar (2003), to Bullier (2001a; 2001b) and to DiLollo et al. (2000), the visual scene would be partially analyzed by rapid and direct projections from the early visual areas to the higher visual areas. This would yield a top-down working hypothesis informing the ongoing ascendant analysis. This descendant modulation reduces the number of possible solutions for a particular retinal stimulation. DiLollo's model is not precisely constrained anatomically. The other two are: In Bar's model, the information is first projected to the prefrontal cortex and comes back to the infero-temporal cortex; and, in Bullier's model, it is sent to V5 and comes back to V1/V2. In both these models the first sweep contains low spatial frequencies. We have already seen empirical evidence for this at the end of the last section (see Bacon et al., 2003). In any case, Bar's and Bullier's models are illustrated in Figure 4.

Insert Figure 4 about here

5.2.1 Empirical evidence for flexible and iterative processing sweeps

Many physiological and electrophysiological studies lend support to the existence of such iterative processes, implicating feedback or reentrant information. We will review the four types of argumentation that have been put forward: 1) the anatomical importance of brain projections (i.e., size of the pathways revealed by tracing and autoradiography techniques) and their functional importance (i.e., revealed by measures of the functional effect of inactivating one cortical area) are not always highly correlated (Vanduffel, Payne, Lomber & Orban, 1997); 2) despite what the topology of the visual system suggests, the visual cortex is temporally compact (Girard, Hupe, & Bullier, 2001; Hupe et al., 2001); 3) the temporal characteristics of neuronal response support the idea that cortical areas are implicated in different visual analyses at different moments in time (Lamme & Roelfsema, 2000; Lee, Mumford, Romero & Lamme, 1998); and 4) low-level cortical areas like V1 can produce sophisticated responses incompatible with their classical function of simple features detectors (Lee et al., 1998). Only the last three points will be detailed next.

5.2.1.1 The visual cortex is temporally compact. Two conditions must be met for the responses of neurons in low-level cortical areas to be modified through time: some neurons in higher-level areas must be activated rapidly; and these areas must feedback rapidly into the low-level cortical areas (Bullier, 2001b). A recent meta-analysis of studies that measured latencies of the visual response of neurons in different cortical areas revealed a temporal hierarchy that diverges considerably from the anatomical (classical) hierarchy (Lamme & Roelfsema, 2000). Furthermore, this meta-analysis showed that neurons in MT and FEF (frontal eye field) areas are activated as rapidly as V1 neurons (MT: minimum = 39ms, mean = 76 ms; FEF: min = 43 ms, mean = 91 ms; V1: min = 35 ms, mean = 72 ms) and are activated more rapidly than neurons located in areas as low as V2 (min = 54 ms, mean = 84 ms) and V3 (min = 50 ms, mean = 77 ms) (Bullier, 2001b; Lamme & Roelfsema, 2000). This is perfectly consistent with Bar's (2003) model as well as with Bullier's (2001a). Numerous factors could contribute to this lack of correspondence between topology and latencies of activation. First, neurons do not receive all their inputs via the shortest possible paths; second, propagation speed of visual information differs according to neuronal pathways: a well-known distinction exists between magnocellular (fast), parvocellular (moderate) and koniocellular (slow) pathways; lastly, it is possible to bypass the LGN through, for example, the superior colliculus and the pulvinar, and to directly feed the extrastriate cortex with visual information (Lamme & Roelfsema, 2000). Concerning the conduction speed of top-down pathways, Girard et al. (2001; see also Panzeri, Rolls, Battaglia & Lavis, 2001) observed fast feedback from V2 to V1 (roughly 3.5 m/s). This speed is most than sufficient to allow for a very rapid influence of high cortical areas on lower ones.

5.2.1.2 A given cortical area is implicated in different analyses at different moments. The response of cortical neurons is not constant. Instead, it seems that cortical neurons participate in different analyses at different moments (Lamme & Roelfsema, 2000). Modulations in neuronal responses across time have already been observed in the LGN (DeAngelis, Ohzawa, & Freeman, 1995), V1 (Ringach, Hawken, & Shapley, 1997, 2003) and IT (Sugase et al., 1999). Ringach et al. (1997), for example, have shown that while the V1 neurons receiving a direct input from LGN (layers 4C et 4C) have a constant preferred orientation through time, the preferred orientation of neurons in subsequent layers (2, 3, 4B, 5 et 6) drastically changes over time. It appears to be impossible to explain modulations such as the one just described with an exclusively feedforward model (Ringach et al., 1997). Instead, Lamme and Roelfsema (2000) proposed that a compulsory feedforward sweep of activation lasting about 100 ms is followed by horizontal (i.e., from the same cortical area) and top-down influence lasting about 200 ms.

5.2.1.3 Low order cortical areas are responsible of sophisticated responses. Cortical neurons, even those of V1, are not simple detectors responding selectively to one particular feature of the visual scene. Some neurophysiological data (e.g. Lee et al., 1998) show that V1 is capable of sophisticated responses comparable with those of Ullman's (1984) visual routines and Marr's (1982) *computations*. As we have just seen, V1 processes different kinds of information over the 40-350 ms post-stimulation period. Although the initial V1 response (40-60 ms) seems to amount to local feature detection, numerous evidences show that subsequent responses (80-200 ms) depend on contextual information and involve higher-level processing (Kosslyn, Thompson, Kim, & Alpert, 1995; Lee et al., 1998). According to Lee et al. (1998), the time-course of the V1 response argues for its gradual involvement in more and more sophisticated computations, and for its implication in tasks as complex as figure-ground segmentation and objects recognition. In sum, V1 would not be a simple module used in the processing of local features but would rather be a high-resolution buffer used for all sorts of visual processing feats.

5.2.2 Deactivation studies.

The most direct evidences for a top-down influence on low-level visual processing perhaps comes from so-called deactivation studies. Hupe et al. (2001) have demonstrated a significant feedback effect of MT on V1, V2 and V3 less than 10 ms after deactivation (see also Girard et al., 2001). A series of recent transcranial magnetic stimulation (TMS) experiments with macaques and humans have shown that top-down processing is necessary for visual consciousness (Pascual-Leone & Walsh, 2001; Ro, Breitmeyer, Burton, Singhal, & Lane, 2003; Walsh & Cowey, 1998). For example, Pascual-Leone and Walsh (2001) impaired the conscious perception of moving phosphenes "normally" produced by stimulating area MT with TMS by stimulating area V1/V2 with TMS, 5 to 40 ms latter. The most plausible interpretation of this result is that MT activation is not sufficient to perceive moving phosphenes, that this perception requires V1/V2 in order to provide a spatial context to the stimulation. Others cortical areas responsible for conscious perception would also suffer from an interruption in top-down communication between MT and V1/V2 by not receiving enough activation (Bullier, 2001a; Pascual-Leone & Walsh, 2001; Pollen, 2003).

6. General Discussion

We could not come to term with the thought of finishing this chapter without having even attempted to incorporate these relatively novel considerations about the necessity of having both compulsory and flexible feature detectors into a unique categorization model. Although we will not describe a fully articulated model here, we will "day dream" about such a model within the SLIP framework.

The question is: What is the optimal way to use a subset of *s* flexible feature detectors to put an unknown object into one or many categories given H_t, a subset of our entire category hierarchy at processing cycle *t*, and C, a set of compulsory feature detectors? Our working idea is to apply Bayes' theorem sequentially (for another example of sequential use of Bayes' theorem in categorization, see Anderson, 1990, 1991) to estimate $P(f_{i,t+1})$, the probability of encountering feature *i* in the environment at processing cycle $t+1$ given all the elements that we have already listed plus $d_{i,t}$, the fact that we have or have not detected this feature either with compulsory or flexible feature at processing cycle *t*. Bayes theorem warrants that

$$P(f_{i,t+1}) = k^{-1}P(f_{i,t})P(d_{i,t}|f_{i,t}) \qquad \text{(Equation 3)},$$

where *k* is equal to $\sum P(f_t)P(d_t|f_t)$. Finally, at processing cycle $t+1$, we shall activate $sP(f_{i,t+1})$ flexible detectors of feature *i*.

To illustrate the computation of the two main components of Equation 3, consider once more Pi Patel on his raft facing an unknown thing. During the first processing cycle, Pi Patel detects *distinctive tiger marks* on the unknown thing with one of its compulsory feature detector. This implies that only the features found in the "tiger" branches of Pi Patel complete category hierarchy should be looked for at time $t+1$; there is absolutely no need to search those found in the "inanimate" branches or in the "all animals except tiger" branches. Fortunately for us, Pi Patel knows only two tigers: "Richard Parker" – defined by the additional *Richard Parker's eye color* feature – and Walt Disney's rendition of "Shere Khan" – defined by the additional *Shere Khan's eye color* feature –, the tiger from Rudyard Kipling's classic. Assuming that both these features are detectable by unique flexible feature detectors, we have $P(f_{i,t}) = .5$, with *i* = {*Richard Parker's eye color*, *Shere Khan's eye color*}. All the other flexible detectors should be given a probability of being part of the unknown object equal to 0. In the Bayes' theorem, this probability function is called the *prior*.

Is that all we can derive from the first processing cycle? No, we can also gain information about what is not out there based on both the flexible and the compulsory feature detectors that were activated but have remained quiet. Suppose for example that some detectors of *Shere Khan's eye color* were activated during this first processing cycle but did not fire. Either the unknown thing does not possess the *Shere Khan's eye color* feature, or it does but the detectors failed to detect it. We have already mentioned the later in the first section of this chapter: in fact, $P(d_{it}|f_{i,t}) = {}_i$. Let us suppose, for the sake of the present illustration, that, in the present case, this quantity is equal to .5. And because no *Richard Parker's eye color* flexible feature detectors were activated during processing cycle *t*, its associated $P(d_{it}|f_{i,t})$ is given a value of 1. This probability function is known as the *likelihood* in Bayes' theorem.

Combining the prior and the likelihood as shown in Equation 3, we obtain probabilities of (.5 * 1) / (.5 * 1 + .5 * .5) = 2/3 for the *Richard Parker's eye color* and (.5 * .5) / (.5 * 1 + .5 * .5) = .1/3 for the *Shere Khan's eye color*. This probability function is called the *posterior* in Bayes' theorem. Finally, the posterior is multiplied by *s* to set the activation level of the flexible feature detectors at processing cycle *t+1*. If Pi Patel could simultaneously activate 90 flexible feature detectors due to limited processing capacities (see section 4.1), he would activate 2/3 * 90 = 60 and 1/3 * 90 = 30, for *Richard Parker's eye color* and for *Shere Khan's eye color*, respectively.

An entirely satisfying account of effective categorization through time would address two more points: 1) how evolutionary pressures promote the cohabitation of mandatory and flexible detectors; and 2) how evolutionary pressures select the fixed detectors. Recent work by Geisler and Diehl (2002, 2003) that combines Bayesian models of perception with Bayesian models of evolution provide a promising research avenue. We intend to fully develop the categorization model outlined above and add such an evolutionary spin to it in the near future.

References

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P. G. & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature, 433*, 68-72.

Anderson, J. R. (1990). *The adaptative character of thought.* New-Jersey: Lawrence Erlbaum Associates, Publishers.

Anderson, J. R. (1991). The adaptative nature of human categorization. *Psychological Review, 98*, 409-429.

Anglin, J. M. (1977). *Word, object and conceptual development.* New York: Norton.

Armony, J.L. and LeDoux, J.E. (2000). How danger is encoded: Towards a systems, cellular, and computational understanding of cognitive-emotional interactions. In M.S. Gazzaniga (Ed.), *The New Cognitive Neurosciences*, 2nd Ed. Cambridge: MIT Press,1067-1079.

Bacon, B. A., Vinette, C., Gosselin, F., Faubert, J. (2003). What primes in unconscious repetition priming. *Journal of Vision*, *3*(9), 234a.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience, 15*(4), 600-609.

Berlin, B. (1992). *Ethnobiological Classification: Principles of Categorization of Plants and Animals in Traditional Societies.* New Jersey: Princeton University Press.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115-117.

Bonnar, L., Gosselin, F. and Schyns, P. G. (2002). Understanding Dali's Slave Market with the Disappearing Bust of Voltaire: A case study in the scale information driving perception. *Perception, 31*, p. 683-691.

Broadbent, D. E. (1958). Perception and communication. New-York: Pergamon Press.

Brown, R. (1958). How shall a thing be called? *Psychological Review, 65*, 14-21.

Bullier, J. (2001a). Feedback connections and conscious vision. *Trends in Cognitive Sciences, 5*(9), 369-370.

Bullier, J. (2001b). Integrated model of visual processing. *Brain Research Reviews, 36*(2-3), 96-107.

Corter, J. E., & Gluck, M. A. (1992). Explaining basic categories: Features predictability and information. *Psychological Bulletin, 111*, 291-303.

Corter, J. E., Gluck, M. A., & Bower, G. H. (1988). Basic levels in hierarchically structured categories. In *Proceedings of the XXth Annual Conference of the Cognitive Science Society* (pp.118-124). New Jersey: Lawrence Erlbaum Associates, Publishers

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1995). Receptive-field dynamics in the central visual pathways. *Trends in Neurosciences, 18*(10), 451-458.

Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J.-F., Poline, J.-B., & Riviere, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience, 4*(7), 752-758.

Delorme, A., Richard, G. & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and human. *Vision Research, 40*, 2187-2200.

Delorme, A. & Thorpe, S. (2001). Face identification using one spike per neuron: resistance to image degradations. *Neural Networks, 14*, 795-803.

Di Lollo, V., Enns, J. T., & Rensick, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129*(4), 481-507.

Estes, W. K. (1994). *Classification and Cognition.* Oxford: Oxford University Press.

Fabre-Thorpe, M., Delorme, A., Marlot, C. & Thorpe, S. (2001). A limit to speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience*, 13, 171-180.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature, 407*, 630-633.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*(1), 1-47.

Geisler, W. S. and Diehl, R. L. (2002). Bayesian natural selection and the evolution of perceptual systems. *Philosophical Transactions of the Royal Society of London, 357*, 419-448.

Geisler, W. S. and Diehl, R. L. (2003). A Bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science, 27*, 379-402.

Gibson, B. M., Wasserman, E. A., Gosselin, F. & Schyns, P. G. (in press). Applying Bubbles to localize features that control pigeons' visual discrimination behavior. *Journal of Experimental Psychology : Animal.*

Girard, P., Hupe, J. M., & Bullier, J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *Journal of Neurophysiology, 85*(3), 1328-1331.

Gosselin, F., Lapalme, J., & McCabe, E. (in preparation). The Sinusoïdal Pulsation of Attention During Categorization.

Gosselin, F., & Schyns, P. G. (1998). Contingency of parts in object concepts. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (p. 1222). New Jersey: Lawrence Erlbaum Associates, Publishers.

Gosselin, F., & Schyns, P. G. (2001a). Bubbles: A technique to reveal the use of information in recognition. *Vision Research, 41*, 2261-2271.

Gosselin, F. & Schyns, P. G. (2001b). Why do we SLIP to the basic-level? Computational constraints and their implementation. *Psychological Review, 108*, 735-758.

Gosselin, F. & Schyns, P. G. (2002). RAP: a new framework for visual categorization. *Trends in Cognitive Science, 6,* 70-77.

Gosselin, F. & Schyns, P. G. (2004). No troubles with Bubbles: A reply to Murray and Gold. *Vision Research*, *44*, 471-477

Gosselin, F. & Schyns, P. G. (in press). A user's guide to Bubbles. *Carnegie Mellon Cognitive Science Series*.

Gosselin, F. & Schyns, P. G. (submitted). Taxonomic knowledge can change the time course of scene perception. *Visual Cognition*.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243-271.

Hoffmann, J., & Ziessler, C. (1983). Objectidentifikation in kunstlichen begriffshierarchien [Object identification in artificial concept hierarchies]. *Zeitschrift fur Psychologie, 194*, 135-167.

Horton, M. S., & Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development, 51*, 708-719.

Humphreys, K., Gosselin, F., Schyns, P. G., Kaufman, J. and Johnson, M. H. Do 7-month-old infants use the same information as adults to process facial identity? *International Conference on Infant Studies*, Toronto, Canada, 2002.

Hupe, J. M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., & Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. *Journal of Neurophysiology, 85*(1), 134-145.

Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General, 126*, 248-277.

Jolicoeur, P., Gluck, M., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 19*, 31-53.

Jones, G. V. (1983). Identifying basic categories. *Psychological Bulletin, 94*, 423-428.

Kersten, D. (1990). Statistical limits to image understanding. In C. Blackermore (Ed.) *Vision: coding and efficiency.* (pp. 32-44). NY: Cambridge University Press.

Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature, 378*(6556), 496-498.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*(11), 571-579.

Lassaline, M. E. (1990). *The Basic Level in Hierarchical Classification.* Unpublished master's thesis, University of Illinois.

Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. *Vision Research, 38*(15-16), 2429-2454.

Li, F. F., VanRullen, R., Koch, C. & Perona, P. (2002)Rapid natural scene categorization in the near absence of attention. *Proceedings of the Natural Academy of Science U.S.A., 99*, 9596-9601.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience, 5*(9), 910-916.

Luce, D. R. (1986). *Response times: their role in infering elementary mental organisation.* Oxford: Oxford University Press.

Mack, A. & Rock, I. (1998). *Inattentional Blindness*. Boston: MIT Press

Markman, E. M. (1989). *Categorization and Naming in Children: Problems of Induction.* Cambridge, Massachusetts: MIT Press.

Markman, E. M., & Hutchinson (1984). Children's sensitivity to constraints on word meanings: Taxonomic vs. thematic relations. *Cognitive Psychology, 16*, 1-27.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.

Mervis, C. B., & Crisafi, M. A. (1982). Order of acquisition of subordinate-, basic-, and superordinate-level categories. *Child Development, 53*, 258-266.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review*, *63*, 81-97

Most, S. B., Scholl, B. J., Clifford, E., & Simons, D. J. (in press). What you see is what you set: Sustained inattentional blindness and the capture of awareness. *Psychological Review*.

Murphy, G. L. (1991). Parts in objects concepts: Experiments with artificial categories. *Memory & Cognition, 19*, 423-438.

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11*, 70-84.

Murphy, G. L., & Smith, E. E. (1982). Basic level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior, 21*, 1-20.

Murray, R. F., & Gold, J. M. (2004) Troubles with bubbles. *Vision Research, 44*, 461-470.

O'Donnell, C., Gosselin, F. & Schyns, P. G. (2002). The acquisition of facial expertise and how that mediates the information utilized to recognize the face. *Journal of Vision, 2(7), 565a*.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34*(1), 72-107.

Panzeri, S., Rolls, E. T., Battaglia, F., & Lavis, R. (2001). Speed of feedforward and recurrent processing in multilayer networks of integrate-and-fire neurons. *Network-Computation in Neural Systems, 12*(4), 423-440.

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science, 292*(5516), 510-512.

Pollen, D. A. (2003). Explicit neural representations, recursive neural networks and conscious visual perception. *Cerebral Cortex, 13*(8), 807-814.

Pothos, E. M., & Chater, N. (1998a). Rational categories. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 848-853). New Jersey: Lawrence Erlbaum Associates, Publishers.

Pothos, E. M. and Chater, N. (1998b). A simplicity principle in human classification. Manuscript submitted for publication.

Reddy, L., Wilken, P. & Koch, C. (2004). Face-gender discrimination is possible in the near absence of attention. *Journal of Vision, 4*, 106-117.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8,* 368--373.

Ringach, D. L., Hawken, M. J., & Shapley, R. (1997). Dynamics of orientation tuning in macaque primary visual cortex. *Nature, 387*(6630), 281-284.

Ringach, D. L., Hawken, M. J., & Shapley, R. (2003). Dynamics of orientation tuning in macaque V1: the role of global and tuned suppression. *Journal of Neurophysiology, 90*(1), 342-352.

Ro, T., Breitmeyer, B., Burton, P., Singhal, N. S., & Lane, D. (2003). Feedback contributions to visual awareness in human occipital cortex. *Current Biology, 13*(12), 1038-1041.

Rosch, E. (1978). Principles of categorisation. In E. Rosch & B. B. Lloyd (Eds.), *Semantic factors in cognition* (pp. 137-168). Hillsdale, NJ: Erlbaum.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-352.

Schyns, P. G., Bonnar, L. & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science, 13*, 402-409.

Schyns, P. G. & Gosselin, F. (2002). A natural bias for basic-level object categorization. *Journal of Vision, 2*(7), 407a.

Schyns, P.G., Jentzsch, I., Johnson, M., Schweinberger, S. R., & Gosselin, F. (2003). A principled method for determining the functionality of ERP components. *Neuroreport, 14*, 1665-1669

Schyns P.G. & Oliva A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations, *Cognition,* 69(3), 243-265.

Sears, C. R. & Pylyshyn, Z. (2000). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology, 54*(1), 1-14.

Simons, D. J. (2000a). Attentional capture and inattentional blindness. *Trends in Cognitive Sciences, 4*, 147-155.

Simons, D. J. (2000b). Current approaches to change blindness. *Visual Cognition, 7*(1-3), 1-15.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*(7), 261-267.

Smith, M., Gosselin, F. & Schyns, P. G., (2004). Receptive fields for flexible face categorizations. *Psychological Science, 15*, 753-761.

Smith, M. L., Gosselin, F., & Schyns, P. G. (submitted). Discrete attentional pulses implement the micro-genesis of face perception.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature, 400*(6747), 869-873.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23*, 457-482.

Thorpe, S., Fize, D. & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520-522.

Tversky, B., & Hemenway, K. (1984). Objects, parts and categories. *Journal of Experimental Psychology: General, 113*, 169-191.

Ullman, S. (1984). Visual routines. *Cognition, 18*(1-3), 97-159.

Vanduffel, W., Payne, B. R., Lomber, S. G., & Orban, G. A. (1997). Functional impact of cerebral connections. *Proceedings of the National Academy of Sciences of the United States of America, 94*(14), 7617-7620.

VanRullen, R., Delorme, A. & Thorpe, S. (2001). Feed-forward contour integration in primary visual cortex based on asynchronous spike propagation. *Neurocomputing, 38-40*, 1003-1009.

VanRullen, R. Gautrais, J., Delorme, A. & Thorpe, S. (1998). Face processing using one spike per neurone. *Biosystems, 48*, 229-239.

VanRullen, R. & Koch, C. (2003a). Competition and selection during visual processing of natural scenes and objects. *Journal of Vision, 3*, 75-85.

Van Rullen, R., & Koch, C. (2003b) Is perception discrete or continuous? *Trends in Cognitive Sciences, 7*, 207-213.

VanRullen, R. & Thorpe, S. (2001a). Is it a bird? Is it a plane? Ultra-rapid visual categorization of natural and artificial objects. *Perception, 30*, 655-668.

VanRullen, R. & Thorpe, S. (2001b). Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Computation, 13*, 1255-1283.

VanRullen, R. & Thorpe, S. (2001c). The time course of visual processing:from early perception to decision-making. *Journal of Cognitive Neuroscience, 13*, 454-461

Vinette, C. and Gosselin, F. (2002). Spatio-temporal use of information in face recognition. *Journal of Vision, 2*(7), 597a.

Vinette, C., Gosselin, F. & Schyns, P. G. (2004). Spatio-temporal dynamics of face recognition in a flash: It's in the eyes! *Cognitive Science*.

Walsh, V., & Cowey, A. (1998). Magnetic stimulation studies of visual cognition. *Trends in Cognitive Sciences, 2*(3), 103-110.

Ward, L. M. (2003). Synchronous neural oscillations and cognitives processes. *Trends in Cognitive Sciences, 7*(12), 553-559.

Wisniewski, E. J., & Murphy, G. L. (1989). Superordinate and basic category names in discourse: a textual analysis. *Discourse Processing, 12*, 245-261.

Figure captions

Figure 1. (a) The four scenes used in this experiment and the corresponding low-level category names learned by all participants ("field", "mountain", "desert", and "dune"), and the high-level categorizations ("flat" and "hilly") that LUMI observers learned. The two histograms illustrate the RT curves of a SLIP putting a scene in a high- and a low-level category. (b) Proportion of the four error types in function of presentation time for each observer group. The solid and dashed lines are, respectively, the average bestfits of the *SLIP* model to the LUMI and to the CHRO participants individual data points. The black curves represent the proportion of errors on *none* of the perceptual dimensions (e.g., respond "field" when presented with a field scene); the green curves represent the proportion of errors on the *lumi*nance dimension (e.g., respond "mountain" when presented with a field scene); the red curves represent the proportion of errors on the *chro*minance dimension (e.g., respond "desert" when presented with a field scene); and the blue curves represent the proportion of errors on *all* dimensions (e.g., respond "dune" when presented with a field).

Figure 2. The leftmost column gives the time-locked classification images extracted from ten subjects using a temporal version of the *Bubbles* technique. Notice the fast oscillation (in the beta bandwidth; a total of about 4 cycles in the temporal window of the experiment) present in all classification images. The rightmost column shows the result of a different analysis of the same data, i.e. information classification images. See main text for details.

Figure 3. (a) This illustrates how Adolphs, Gosselin, Buchanan, Tranel, Schyns and Damasio (2005) sampled their faces at five independent bands of spatial frequencies (one octave each, with cutoffs at 22.38, 11.19, 5.59, 2.80, and 1.40 cycles per deg), to yield sparsely revealed images whose integration resulted in the final stimuli that subjects saw (far bottom right). (b) Average difference between the effective use of information by SM and the 10 normal controls. (c) Results of Bacon, Vinette, Gosselin and Faubert (2003). The top row contains the diagnostic regions for the conscious condition (i.e., high-spatial frequencies) and the bottom row contains the diagnostic information for the unconscious condition (i.e., low-spatial frequencies).

Figure 4. This outlines the models of Bullier (2001a) and of Bar (2003). According to Bullier (2001a), an iterative loop joins the V1/V2 area and MT/MST one. According to Bar (2003), an analogue loop starts from the V2/V4 area, goes through ventrolateral and orbital PFC and retro-injects information into IT and the amygdala. (LGN : lateral geniculate nucleus; PFC : prefrontal cortex; IT : inferotemporal).

Figure 1.

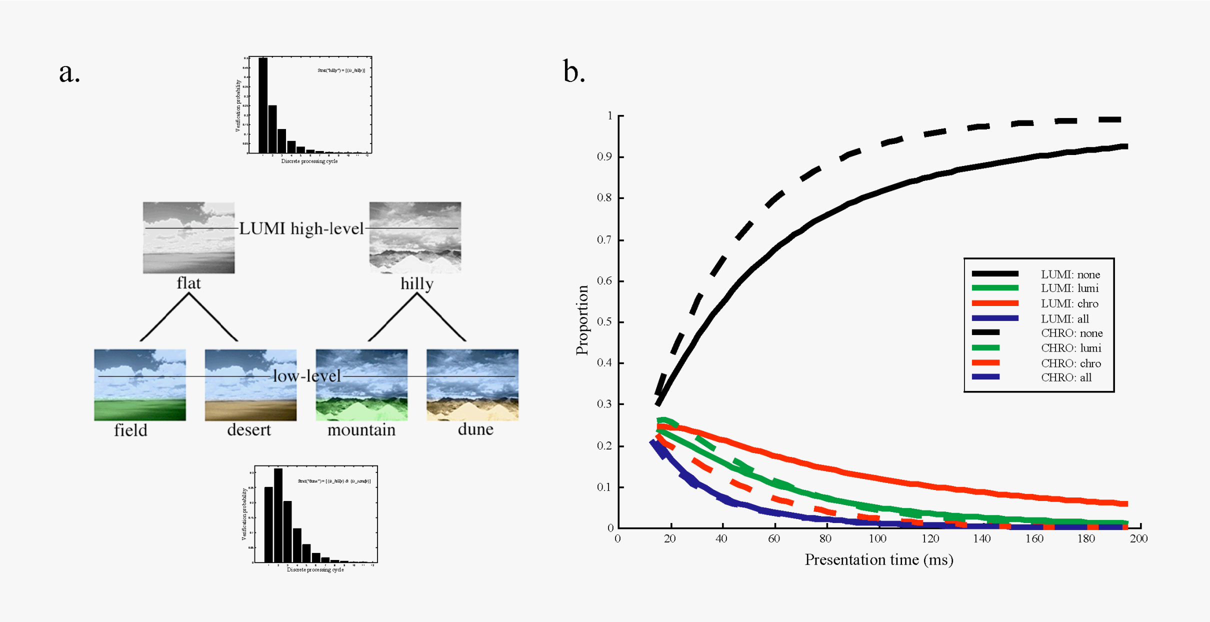
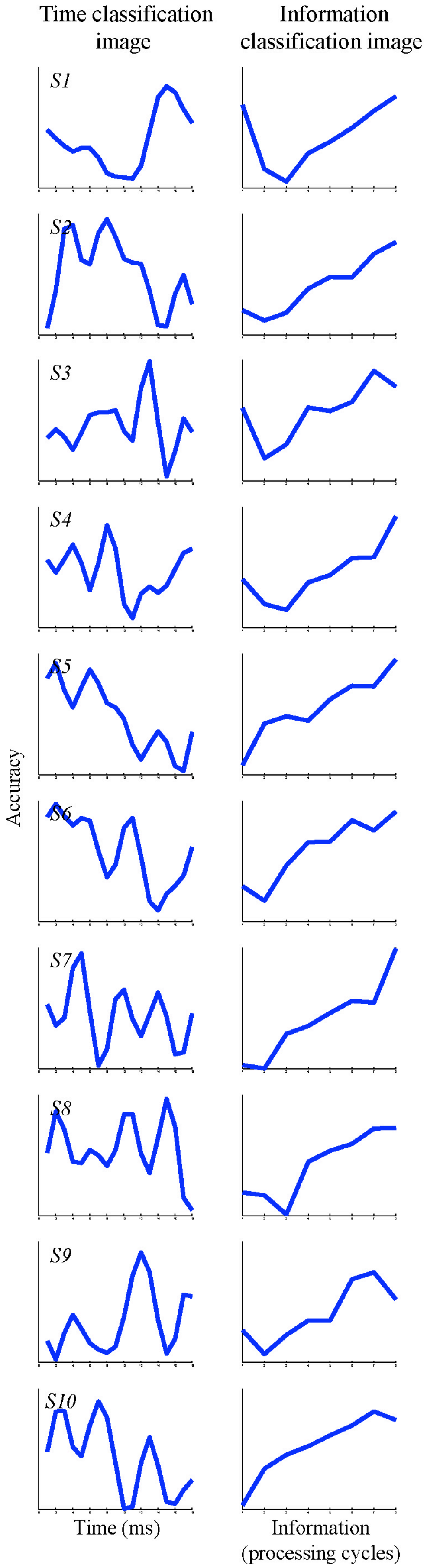

a.
Strat("hilly") = [{is_hilly}]
Verification probability
Discrete processing cycle
LUMI high-level
flat
hilly
low-level
field
desert
mountain
dune
Strat("dune") = [{is_hilly} & {is_sandy}]
Verification probability
Discrete processing cycle
b.
Proportion
Presentation time (ms)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
20
40
60
80
100
120
140
160
180
200
LUMI: none
LUMI: lumi
LUMI: chro
LUMI: all
CHRO: none
CHRO: lumi
CHRO: chro
CHRO: all

Figure 2.



Time classification image

Information classification image

Figure 3.

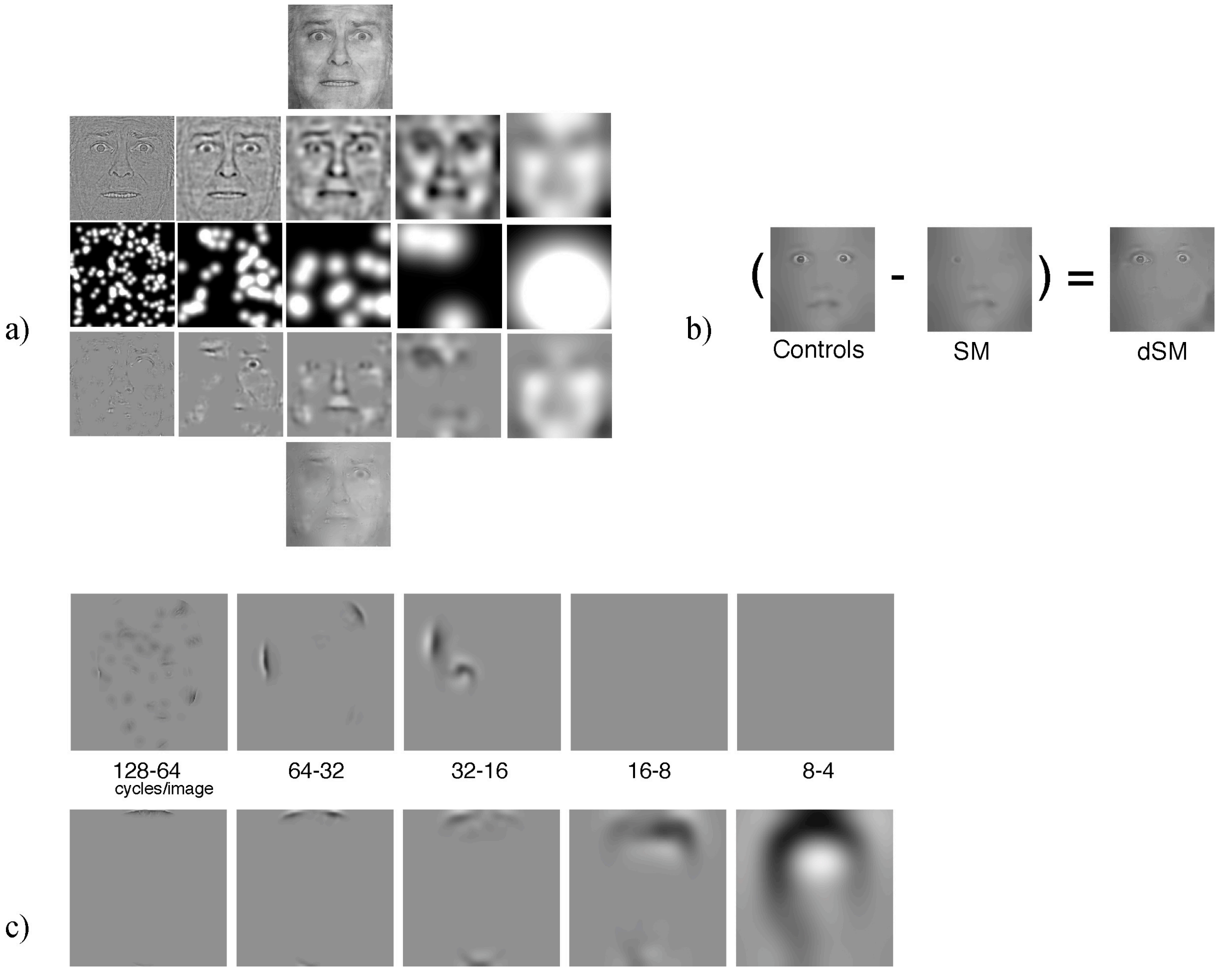
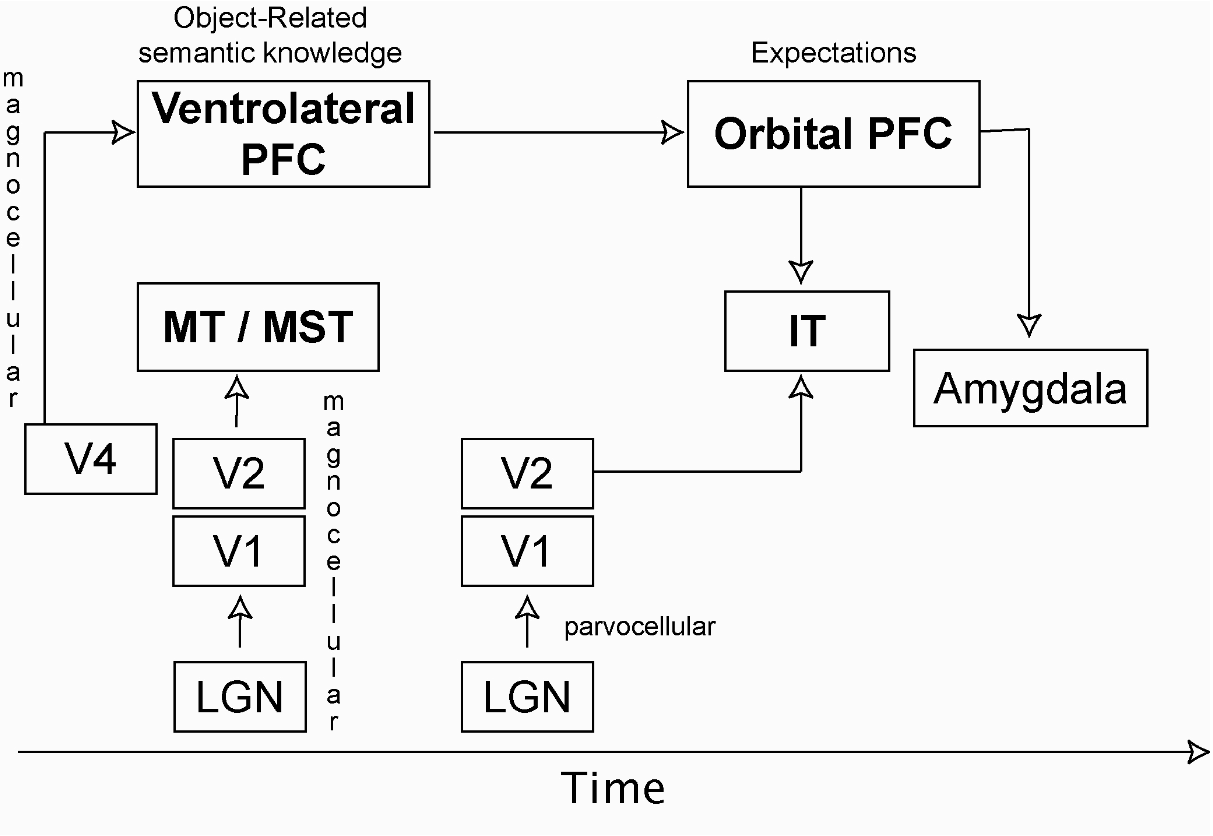


Figure 4.


Object-Related
semantic knowledge
Ventrolateral PFC
Expectations
Orbital PFC
magnocellular
MT / MST
IT
Amygdala
V4
V2
V1
LGN
magnocellular
V2
V1
LGN
parvocellular
Time