

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/317510319>

# The Time Course of Object, Scene, and Face Categorization

Chapter · June 2017

DOI: 10.1016/b978-0-08-101107-2.00036-1

CITATIONS

0

READS

55

4 authors, including:



**Laurent Caplette**

Université de Montréal

16 PUBLICATIONS 8 CITATIONS

[SEE PROFILE](#)



**Caroline Blais**

Université du Québec en Outaouais

88 PUBLICATIONS 707 CITATIONS

[SEE PROFILE](#)



**Frédéric Gosselin**

Université de Montréal

210 PUBLICATIONS 4,942 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Top-down modulation of spatial frequency extraction [View project](#)



Visual information in the brain: linking extraction and processing [View project](#)

All content following this page was uploaded by [Laurent Caplette](#) on 15 September 2017.

The user has requested enhancement of the downloaded file.

**The Time Course of Object, Scene and Face Categorization**

Laurent Caplette<sup>1</sup>, Éric McCabe<sup>1</sup>, Caroline Blais<sup>2</sup> & Frédéric Gosselin<sup>1</sup>

[This is not the final version; slight modifications may have been made afterwards.]

<sup>1</sup>Département de psychologie, Université de Montréal, Montréal, QC, Canada

<sup>2</sup>Département de psychologie et de psychoéducation, Université du Québec en Outaouais,  
Gatineau, QC, Canada

Correspondence information

Laurent Caplette: laurent.caplette@umontreal.ca, 1-514-343-7550

Éric McCabe: eric.mc.cabe@umontreal.ca, 1-514-343-7550

Caroline Blais: caroline.blais@uqo.ca, 1-819-595-3900

Frédéric Gosselin: frederic.gosselin@umontreal.ca, 1-514-343-7550

### Abstract

We first describe *SLIP* (Strategy Length & Internal Practicability), a formal model for thinking about categorization, in particular about the time course of categorization. We then discuss an early application of this model to basic-levelness. We then turn to aspects of the time course of categorization 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. We propose a two-fold resolution of this paradox, attempting, in the process, to bridge work done on categorization in vision, neuropsychology and physiology.

Keywords: categorization; attention; vision; temporal processing; object recognition; scene recognition; face recognition.

### 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 a more abstract level of categorization these stimuli can be categorized as “flat” and “hilly” (Figure 1). Thus, the “mountain” and the “dune” scenes are “hilly”, and the “field” and “desert” scenes are “flat”. We thus have a small category hierarchy. And we are ready to begin the unpacking of our ideas about the time course of categorization.

-----  
 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; Gershman, Horvitz, & Tenenbaum, 2015). A strategy comprises sets of noisy detectors. For example, the dune scene illustrated in Figure 1 satisfies two category strategies:  $\text{Strat}(\text{“dune”}) = [\{is\_hilly\} \& \{is\_sandy\}]$ , is the SLIP strategy for the “dune” category and comprises two sets of detectors;  $\text{Strat}(\text{“hilly”}) = [\{is\_hilly\}]$  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, only one successful luminance detector is sufficient to verify that a scene is “flat” in this category hierarchy. 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 categorize a scene. For example, to verify that a scene is a “dune” in the category hierarchy, one successful luminance detector and one successful chrominance detector are required. Everything else being equal, SLIP predicts that 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:

$$\mu = \prod_{j=1}^n (1 - \phi_j^t) \quad (\text{Equation 1}),$$

with  $\phi_j = 1 - \gamma_j$ . The constant  $\gamma_j$  is the probability that the set of detectors  $j$  is successful after one processing cycle; it constitutes the “weight” given to dimension  $j$  (for details see Gosselin & Schyns, 2001a).

Although the probability distribution of Equation 1 is useful in its own right, we have more often employed its associated density function. This is the function that determines 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:

$$\chi = \mu(t) - \mu(t-1) \quad (\text{Equation 2}).$$

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  $\phi$ ’s (i.e., the probability that the set of detectors  $j$  is unsuccessful during one processing cycle) were equal to .5. Two properties of these density functions are noteworthy: (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  $\sum_{t=1}^{+\infty} t\chi(t)$ , i.e. 2 cycles for Strat(“hilly”) and 2.67 cycles for Strat(“dune”); (2) Another noteworthy aspect of these density functions is their shapes which are reminiscent of response time (RT) density functions (e.g., Luce, 1986). If we assume that 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 significant portion 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). This is the first of many verification experiments that demonstrated a superiority at basic level of abstraction.

But this basic-level superiority doesn’t always hold. An exception identified early on in the literature is that of atypical exemplars of a basic-level category that are categorized faster at the subordinate than at the basic level (e.g., a penguin is categorized as a "penguin" faster than as a "bird"; Murphy & Brownell, 1985). Another exception is that of dog and bird enthusiasts that are faster at verifying the objects of their enthusiasm at the subordinate than at the basic category (e.g., a Doberman is categorized faster as a "Doberman" than as a "dog" by a dog enthusiast; Tanaka & Taylor, 1991). More recently, Macé et al. (2009) have observed faster "ultra-rapid categorization" at the superordinate level than at the basic-level (more on the subject in section *4.1 Compulsory feedforward processing sweeps*), and suggested that, “You spot the animal faster than the bird.” Mack and Palmeri (2015) examined some of the requirements for this ultra-rapid categorization superordinate level superiority. They showed that with randomized target categories, the superordinate advantage was eliminated; and with only four repetitions of superordinate categorization within an otherwise randomized context, the basic-level advantage was eliminated.

The SLIP model is perfectly adapted to predicting basic-levelness, i.e., the average speed of categorization at various levels of abstraction in a verification task, including the cases that

favor levels of abstraction other than basic. Gosselin and Schyns (2001a) compared the predictive power of SLIP with that of four other basic-levelness 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, 2002). To do so, they drew 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 from three novel experiments using computer-synthesized 3-D artificial objects. SLIP led the pack by a large margin, predicting 88% of this data set, while the category utility model, the category feature-possession model the compression measure, and the context model predicted, respectively, 64%, 62%, 42% and 35% of this data set.

### *3. The need for flexibility and a paradox*

So far our story has been, to a large extent at least, a success story. We now turn to limitations of the SLIP model and of other similar models, and to ways to overcome these limitations. We hope that this will point toward new directions for research on the time course of categorization.

#### *3.1 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; see Cowan, 2000, for a critical review); 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., Most, Scholl, Clifford & Simons, 2005; Rensink, O’Regan & Clark, 1997; Simons & Levin, 1997; Simons, 2000b, see the special issue of Visual Cognition). In the related paradigm of “inattention 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 a team in white T-shirts pass a basketball). In a 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). Typically less than one fourth of the participants notice these odd events (Mack & Rock, 1998; Simons, 2000a).

As we have written in the first section of this chapter, the SLIP model can implement this psychological reality by having only a subset of all available detectors activated simultaneously. Thus the blindness to a large change occurring in a natural scene – say an engine disappearing from the wing of an airplane – could result from not enough of the relevant SLIP detectors being active simultaneously for the verification to be completed on time. Increasing the number of active detectors would result in the change being detected.

### *3.2 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 evidence for this in section 4.2 *Flexible iterative processing sweeps*. As a preview, here we will consider an experiment performed with the four artificial scenes of Figure 1. In a learning phase, all participants learned to categorize the four scenes at a general and at a specific level. Participants were split in two groups (called LUMI and CHRO) that differed on the categories at the 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 (2000) 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<sup>1</sup>. The insight is that each group is tuned to chromatic and luminance information to maximize their categorization potential. After a successful test only on the luminance dimension, the LUMI group can already categorize the scene at a general level — put the dune scene in the “hilly”

---

<sup>1</sup> Gosselin (2000) adapted SLIP to predicting the error patterns of subjects in this situation. In a nutshell, he 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.

category, for example — whereas the CHRO group cannot. In contrast, after a successful test only on the chrominance dimension, the CHRO group can categorize the scene at a general level — put the dune scene in the “sandy” category — whereas the LUMI group cannot.

Here, this does not help the participants but, in real life, putting an object in a category, even when it’s not the target category, allows them to quickly 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 lapse of time: “This is Richard Parker”. However, the initial testing of the *tiger marks* in the first route could lead to an intermediate “tiger” categorization before the specific “Richard Parker” categorization, allowing Pi Patel to react faster.

Very little is known about how categorizations are embedded in real life. Gosselin’s 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 macaques. 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. The same seems to apply to object recognition. In a behavioral recognition task with high temporal resolution, De la Rosa, Choudhery and Chatziastros (2011) observed that object detection occurred earlier than categorization and identification.

### *3.3 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 others 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.

## *4. Categorization as an iterative process*

Our resolution of the above paradox will be two-fold: We will argue that a subset of detectors of a SLIP categorizer – or any other categorizer – 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.

### *4.1 Compulsory feedforward processing sweeps*

Ullman (1984) was among the first to propose that object recognition is informed by feedforward compulsory processing sweeps. 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.

Several studies performed by Thorpe and his research group as well as by others demonstrate the capacity of human subjects to categorize a 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, 2001b). 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, 2003): the performance of subjects in such tasks does not increase with training even for a 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 feedforward manner by the amygdala and can subsequently drive behaviour. One question that remains largely unanswered by all these experiments concerns the nature of the information processed during these compulsory feedforward sweeps.

#### *4.1.1 The nature of the information processed during compulsory feedforward sweeps*

Caplette et al. (submitted) investigated the nature of the information sampled across time when categorizing an object. They used the *Bubbles* technique to uncover which spatial frequencies shown at which moments correlate with accurate object recognition in neurotypical individuals and subjects with Autism Spectrum Disorder (ASD). We describe this experiment in some details in the next section, which can also be read as a *Bubbles* primer.

*4.1.1.1 A Bubbles Primer.* “*Bubbles*” is a generic procedure that can reveal the information that drives a measurable response (Gosselin and Schyns, 2001b). Six decisions, or answers to six questions, are required in order to set up a *Bubbles* experiment (Gosselin and Schyns, 2005): (1) what is the stimulus set? (2) in which space will the 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) what are the observers? Next, we discuss each of these decisions in the context of Caplette et al. (submitted).

1) *Stimulus Set?* In a *Bubbles* experiment, the stimulus set is crucial, because it critically bounds what will be tested. Caplette et al. used 86 object images. These images were gray-shaded and had a resolution of 256 x 256 pixels (subtending 6 x 6° of visual angle). The overall energy and the spatial frequency content of the images were normalized. The images were presented for a duration of 333 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 that are atypical of natural processing.

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. For instance, stimuli used in a *Bubbles* experiment have been sampled in Cartesian space (e.g., Gosselin & Schyns, 2001b; Gibson et al., 2005; 2007; Rutishauser et al., 2013), in Fourier space (e.g., Willenbockel et al., 2010; Caplette et al., 2014), in time (Blais, Arguin & Gosselin, 2013), and in Cartesian x time space (e.g., Vinette, Gosselin & Schyns, 2004; Blais et al., 2009). Caplette et al. searched the temporal and spatial frequency dimensions with a resolution of 120 Hz and 0.5 cycle per image (cpi).

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, spatial frequencies were randomly sampled through time. Sampling matrices were created by convolving sparse

matrices of binary noise with a 2-D Gaussian kernel (i.e., a bubble). Accuracy was maintained at 75% correct by adjusting the number of bubbles on a trial-by-trial basis using a gradient descent algorithm.

4) *The Task?* At this stage, the sampling procedure has been fully specified. The next decision is that of the task. In Caplette et al., it was a verification task. After each dynamic bubbled stimulus, an object name was shown and subjects had to indicate as rapidly and as accurately as possible if the name matched the object (it did on 50% of the trials). Several tasks can and have been used with the same stimulus set.

5) *The Response?* The response is a critical parameter of a *Bubbles* experiment because it allows to target different encapsulated mechanisms and even different processing stages of these mechanisms. In Caplette et al., observers pressed labeled keys to indicate if the name matched the object or not. Such key-press responses have been used to derive correct and incorrect responses (e.g., Gosselin & Schyns, 2001b; Adolphs et al., 2005; Vinette, Gosselin & Schyns, 2004; Gibson et al., 2007; Fiset et al., 2008; Thurman et Grossman, 2008; Langner, Becker et Rinck, 2009; Willenbockel et al., 2010; Lee et al., 2011; Blais et al., 2012; Tadros et al., 2013; Roy et al., 2015), and response latencies (Caplette et al., 2014; Dupuis-Roy et al., 2009; Schyns, Bonnar & Gosselin, 2002; Willenbockel et al., 2010). Past experiments have also used electroencephalographic (EEG) activity (e.g., Schyns et al., 2003; Smith, Gosselin & Schyns, 2004, 2006, 2007), magnetoencephalography (MEG; Smith et al., 2009), functional magnetic resonance imaging (fMRI; Smith et al., 2008), and firing rate of single cells (e.g., Wang et al., 2014; Rutishauser et al., 2013). Other responses could be galvanic skin response, plethysmographic response, pupillometric response, and so forth. To the extent that *Bubbles* is essentially an empirical tool, it is recommended to record as many different responses as possible

(e.g., correct/incorrect, latencies and EEG in a face recognition experiment). It is sometimes difficult to predict before the experiment how responses will correlate with the parameters of the search space.

6) *The Observers?* Depending on the objectives of the research, different types of observers can interact with the bubbled stimuli. Caplette et al. used human observers — neurotypical and ASD subjects, but only the neurotypical subjects will be discussed here. Brain lesion patients (Adolphs et al., 2005; Caldara et al., 2005), psychiatric patients (Lee et al., 2011; Clark, Gosselin & Goghari, 2013), animals (e.g., Gibson et al., 2005; 2007; Nielsen, Logothetis & Rainer, 2006a; 2006b; Vermaercke & Op de Beeck, 2012) and model observers (e.g., Gosselin & Schyns, 2001b) have also performed *Bubbles* experiments.

4.1.1.2 *A coarse-to-fine sampling.* Once 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 (e.g., Chauvin et al., 2005; Murray, 2012). In Caplette et al., this results in an image giving us the correlation between the presentation of a certain spatial frequency in a certain temporal slot and accurate responses, i.e. a time x spatial frequency *classification image*. Figure 2 shows the random-effects classification image of forty-nine neurotypical subjects that each completed 2,000 trials.

A visual inspection reveals for the first time that neurotypicals sample lower spatial frequencies continuously and gradually sample higher and higher spatial frequencies. Previous studies examining spatial frequency sampling across time (Hughes, Nozawa, & Kitterle, 1996; Loftus & Harley, 2004; Parker, Lishman, & Hughes, 1992; 1997; Schyns & Oliva, 1994) had compared a handful of spatial frequency x time conditions —two typically — and thus lacked the

resolution needed to reveal this information sampling pattern. We suspect that the low spatial frequencies are sampled continuously to allow us to discover unexpected objects and flexibly activate the detectors required to inspect them (see section *4.1 Compulsory feedforward processing sweeps*).

-----  
Insert Figure 2 about here  
-----

#### *4.2 Flexible iterative processing sweeps*

In light of the evidence presented in the previous section, we can say that some information is always being sampled or, using SLIP terminology, some detectors are active throughout all processing cycles. This information appears to be a band of relatively low spatial frequencies from 1 to approximately 20 cpi. Higher and higher spatial frequencies are sampled afterwards. This is a kind of coarse-to-fine visual routine. An important question remains however: is this routine ballistic or adaptive? That is, are later processing sweeps modulated by the initial feedforward processing sweep?

##### *4.2.1 Ballistic vs adaptive visual routines*

That we have uncovered significant blobs in our time x spatial frequency linear classification image indicates that our sampling of visual information is at least partly ballistic: some information is better processed at specific moments, independently of what was sampled before (see also Loftus & Harley, 2004). However, our visual routine can still be adaptive to some extent. For example, our visual system might always sample high spatial frequencies in later processing cycles but do so more efficiently if low spatial frequencies were sampled

beforehand. Moreover, sampling could be ballistic on the spatial frequency dimension while being adaptive on another dimension. For instance, low spatial frequencies might inform us about which features at specific locations in the visual field to sample in high spatial frequencies afterwards. Many neuroscientists have argued that rapidly extracted coarse information modulates top-down the extraction of finer information (e.g., Bar, 2003; Bullier, 2001b; DiLollo et al., 2000; Hochstein & Ahissar, 2002).

#### 4.2.2 *Top-down models of object recognition*

Compulsory feedforward sweeps fit 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 3.2 *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 three models that mix early compulsory feedforward 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 (Caplette et al., submitted). Further evidence supports the existence of top-down processes during object recognition, and the role of low spatial frequencies in them; we discuss these in the next subsections. In any case, Bar's and Bullier's models are illustrated in Figure 3.

-----  
Insert Figure 3 about here  
-----

#### *4.2.3 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 first review three types of indirect evidence that have been put forward: despite what the topology of the visual system suggests, the visual cortex is temporally compact (Girard, Hupé, & Bullier, 2001; Hupé et al., 2001); 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 low-level cortical areas like V1 can produce sophisticated responses incompatible with their classical function of simple feature detectors (Lee et al., 1998). We will then review more direct evidence that there is feedback from high order areas onto early visual areas during object recognition, and that low spatial frequencies play a particular role in the initiation of this feedback.

*4.2.3.1 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 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 = 39 ms, 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 (2001b). 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 more than sufficient to allow for a very rapid influence of high cortical areas on lower ones.

*4.2.3.2 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 $\alpha$  et 4C $\beta$ ) 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 (see also Mazer, Vinje, McDermott, Schiller, & Gallant, 2002). 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.

*4.2.3.3 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 object 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.

*4.2.3.4 High order areas feedback on early visual areas.* Some studies provide more direct evidence of top-down mechanisms in vision. Hupé 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), and a series of 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 later. 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. Other 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).

*4.2.3.5 Low spatial frequencies initiate top-down processing.* In an object recognition task in MEG and fMRI, Bar et al. (2006) revealed distinctive activity in the orbitofrontal (OFC) cortex when an object was recognized, 50 ms earlier than in recognition-related areas in the temporal cortex. Moreover, this activity was preferentially modulated by low spatial frequencies (see also Kauffman, Bourgin, Guyader, & Peyrin, 2015). Results from Kveraga, Boshyan, & Bar (2007) using stimuli biased toward magnocellular processing in luminance and color further suggested that fast magnocellular projections (known to transmit lower SFs) were playing a critical role in the initiation of a top-down processing from the OFC. Peyrin et al. (2010) also demonstrated that higher frontal and temporal areas responded more to low SF stimuli when they

were presented before high SF stimuli, and that occipital areas responded more to high SF stimuli when they were presented after low SF stimuli. This suggests that low SFs initiate a top-down processing from higher areas, which enhances neural activity to high SFs in the early visual cortex. It has been shown that top-down expectations elicit a template of the stimulus representation in early visual areas (Kok, Jehee, & de Lange, 2012; Kok, Failing, & de Lange, 2014): hence, this enhanced neural activity is likely to be a coarse representation of the stimulus (and not only an unspecific global attention effect). Recently, Goddard et al. (2016) did observe, using Granger causality analysis, an early feedback flow of information driven by low SFs. Together, these findings support the initiation of top-down mechanisms, or flexible iterative processing sweeps, by low spatial frequencies during object recognition.

### *5. General Discussion*

We could not come to terms with 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 categorization model. Although we will not describe a fully articulated model here, we will “day dream” about such a model within the SLIP framework.

So 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; Gershman, Horvitz, & Tenenbaum, 2015) 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

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

where  $k$  is equal to  $\sum_i P(f_{i,t})P(d_{i,t}|f_{i,t})$ . And, finally, at processing cycle  $t+1$ , to 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  $i = \{\text{Richard Parker’s eye color, Shere Khan’s eye color}\}$ . 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 latter in the first section of this chapter: in fact,  $P(d_{ii}|f_{i,t}) = \phi_i$ . 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  $P(d_{ii}|f_{i,t})$  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  $t+1$ . Let's assume for the sake of the argument that Pi Patel can simultaneously activate 90 flexible feature detectors (see section 4.1), he would thus 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 the time course of object, scene and face categorization would address two more points: 1) how do evolutionary pressures promote the cohabitation of mandatory and flexible detectors; and 2) how do evolutionary pressures select the fixed detectors. Work by Geisler and Diehl (2002, 2003) that combines Bayesian models of perception with Bayesian models of evolution seems like the most promising avenue to tackle these issues.

We have presented *SLIP*, a formal model for thinking about the time course of object, scene and face categorization. Next, we argued that our limited visual processing capacities lead to the necessity of having a flexible categorization apparatus, and that this inexorably brings about a paradox — how can we flexibly adjust to the environment before knowing what is out there? Finally, we proposed an embryonic solution of this paradox within the *SLIP*

framework, sampling ideas from the fields of vision, neuropsychology and physiology. That is, detectors of low spatial frequencies might be active throughout all recognition to allow unexpected objects to be discovered and flexibly guide the activation of other detectors (i.e. determine which features to extract afterwards). We have provided some evidence toward this possibility; however, the existence of such an adaptive process remains to be directly investigated.

## 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 adaptive character of thought*. New-Jersey: Lawrence Erlbaum Associates, Publishers.
- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409-429.
- 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.
- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600-609.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 449-454.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115-117.
- Blais, C., Arguin, M., & Gosselin, F. (2013). Human visual processing oscillates: evidence from a classification image technique. *Cognition*, *128*(3), 353-362.
- Blais, C., Fiset, D., Arguin, M., Jolicœur, P., Bub, D., & Gosselin, F. (2009). Reading between Eye Saccades. *PLoS One*, *4*(7), e6448-7.
- Blais, C., Roy, C., Fiset, D., Arguin, M., & Gosselin, F. (2012). The eyes are not the window to basic emotions. *Neuropsychologia*, *50*(12), 2830-2838.
- Broadbent, D. E. (1958). *Perception and communication*. New-York: Pergamon Press.

- 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.
- Caplette, L., West, G., Gomot, M., Gosselin, F., & Wicker, B. (2014). Affective and Contextual Values Modulate Spatial Frequency Use in Object Recognition. *Frontiers in Psychology*, 5:512.
- Caplette, L., Wicker, B., & Gosselin, F. (submitted). Atypical Time Course of Object Recognition in Autism Spectrum Disorder.
- Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, 5(9), 659–667.
- Clark, C. M., Gosselin, F. & Goghari, V. M. (2013). Aberrant patterns of visual facial information usage in schizophrenia. *Journal of Abnormal Psychology*, 122, 513–519.
- 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
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity, *Behavioral and Brain Sciences*, 24(1), 87–185.
- DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1995). Receptive-field dynamics in the central visual pathways. *Trends in Neurosciences*, 18(10), 451-458.
- De la Rosa, de, S., Choudhery, R. N., & Chatziastros, A. (2011). Visual object detection, categorization, and identification tasks are associated with different time courses and

- sensitivities. *Journal of Experimental Psychology. Human Perception and Performance*, 37(1), 38–47.
- 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.
- Fiset, D., Blais, C., Éthier-Majcher, C., Arguin, M., Bub, D. & Gosselin, F. (2008). Features for uppercase and lowercase letter identification. *Psychological Science*, 19, 1161-1168.
- 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.

- Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273-278.
- Gibson, B. M., Wasserman, E. A., Gosselin, F. & Schyns, P. G. (2005). Applying Bubbles to localize features that control pigeons' visual discrimination behavior. *Journal of Experimental Psychology : Animal*.
- Gibson, B., Lazareva, O. F., Gosselin, F., Schyns, P. G. & Wasserman, E. A. (2007). Non-accidental properties underlie shape recognition in mammalian and non-mammalian vision. *Current Biology*, *17*, 336-340.
- Girard, P., Hupé, 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.
- Goddard, E., Carlson, T. A., Dermody, N., & Woolgar, A. (2016) Representational dynamics of object recognition: Feedforward and feedback information flows. *Neuroimage*, *128*, 385–397.
- 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. (2000). Why do we SLIP to the basic-level? A formal model. Unpublished doctoral thesis, University of Glasgow.
- Gosselin, F. & Schyns, P. G. (2001a). Why do we SLIP to the basic-level? Computational constraints and their implementation. *Psychological Review*, *108*, 735-758.
- Gosselin, F., & Schyns, P. G. (2001b). Bubbles: A technique to reveal the use of information in recognition. *Vision Research*, *41*, 2261-2271.

- Gosselin, F. & Schyns, P. G. (2005). Bubbles: A user's guide. In L. Gershkoff-Stowe & D. H. Rakison (Eds.) *Building Object Categories in Developmental Time* (pp. 91-106). Hillsdale, NJ: Erlbaum.
- Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804.
- Hoffmann, J., & Ziessler, C. (1983). Objectidentifikation in kunstlichen begriffshierarchien [Object identification in artificial concept hierarchies]. *Zeitschrift fur Psychologie*, *194*, 135-167.
- Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, *8*(3), 197–230.
- Hupé, 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.
- Kauffmann, L., Bourgin, J., Guyader, N., & Peyrin, C. (2015). The Neural Bases of the Semantic Interference of Spatial Frequency-based Information in Scenes. *Journal of Cognitive Neuroscience*, *27*(12), 2394-2405.
- 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.
- Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in

- the primary visual cortex. *Journal of Cognitive Neuroscience*, 26(7), 1546–1554.
- Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, 75(2), 265–270.
- 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. 1114.
- Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience*, 27(48), 13232–13240.
- 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.
- Langner, O., Becker, E. S., & Rinck, M. (2009). Social Anxiety and Anger Identification Bubbles Reveal Differential Use of Facial Information With Low Spatial Frequencies. *Psychological Science*, 20(6), 666-670.
- Lassaline, M. E. (1990). *The Basic Level in Hierarchical Classification*. Unpublished master's thesis, University of Illinois.
- Lee, J., Gosselin, F., Wynn, J. K., & Green, M. F. (2011). How do schizophrenia patients use visual information for decoding facial emotion? *Schizophrenia Bulletin*, 37, 1001-1008.
- 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.

- Loftus, G. R., & Harley, E. M. (2004). How different spatial-frequency components contribute to visual information acquisition. *Journal of Experimental Psychology. Human Perception and Performance*, 30(1), 104–118.
- Macé, M. J.-M., Joubert, O. R., Nespoulous, J.-L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations: You spot the animal faster than the bird. *PLoS ONE*, 4(6), e5927.
- Mack, A. & Rock, I. (1998). *Inattentional Blindness*. Boston: MIT Press
- Mack, M. L., & Palmeri, T. J. (2015). The Dynamics of Categorization: Unraveling Rapid Categorization. *Journal of Experimental Psychology : General*, 144, 3, 551-569.
- Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.
- Mazer, J. A., Vinje, W. E., McDermott, J., Schiller, P. H., & Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3), 1645–1650.
- 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. R., & Simons, D. J. (2005). What you see is what you set: sustained inattention blindness and the capture of awareness. *Psychological review*, 112(1), 217.
- 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 RF (2012). Classification images and bubbles images in the generalized linear model. *Journal of Vision*, *12*(7):2, 1-8.
- Nielsen, K. J., Logothetis, N. K., & Rainer, G. (2006a). Discrimination strategies of humans and rhesus monkeys for complex visual displays. *Current Biology*, *16*(8), 814-820.
- Nielsen, K. J., Logothetis, N. K., & Rainer, G. (2006b). Dissociation between local field potentials and spiking activity in macaque inferior temporal cortex reveals diagnosticity-based encoding of complex objects. *The Journal of neuroscience*, *26*(38), 9639-9645.
- 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.
- Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, *21*, 147-160. doi:10.1068/p210147
- Parker, D. M., Lishman, J. R., & Hughes, J. (1997). Evidence for the view that temporospatial integration in vision is temporally anisotropic. *Perception*, *26*, 1169-1180. doi:10.1068/p26116
- 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.
- Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., et al. (2010). The neural substrates and timing of top-down processes during coarse-to-fine categorization of

- visual scenes: a combined fMRI and ERP study. *Journal of Cognitive Neuroscience*, 22(12), 2768–2780.
- 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. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science*, 26(3), 303-343.
- 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., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-352.
- Roy, C., Blais, C., Fiset, D. Rainville, P., & Gosselin, F. (2015). Efficient information for recognizing pain in facial expressions. *European Journal of Pain*, 19, 852-860.

- Rutishauser, U., Tudusciuc, O., Wang, S., Mamelak, A. N., Ross, I. B., & Adolphs, R. (2013). Single-neuron correlates of atypical face processing in autism. *Neuron*, *80*(4), 887-899.
- 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., 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. (1994). From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, *5*(4), 195–200.
- 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 inattention 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. (2006). Perceptual moments of conscious visual experience inferred from oscillatory brain activity. *Proceedings of the National Academy of Sciences*, *103*, 5626-5631.
- Smith, M. L., Gosselin, F. & Schyns, P. G. (2007). From a face to its category via a few information processing states in the brain. *NeuroImage*, *37*, 974-984.

- Smith, M. L., Fries, P., Goebel, R., Gosselin, F., & Schyns, P. G. (2009). Feature-driven decomposition of brain measurements—a novel approach to interpreting MEG brain activity during categorization tasks. *Cerebral Cortex, 19*, 2428-2438.
- 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.
- Tadros, K., Dupuis-Roy, N., Fiset, D., Arguin, M., & Gosselin, F. (2013). Reading laterally: the cerebral hemispheric use of spatial frequencies in visual word recognition. *Journal of Vision, 13*(1):4.
- 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.
- Thurman, S. M., & Grossman, E. D. (2008). Temporal “Bubbles” reveal key features for point-light biological motion perception. *Journal of Vision, 8*(3), 28.
- Ullman, S. (1984). Visual routines. *Cognition, 18*(1-3), 97-159.
- 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. (2003). Competition and selection during visual processing of natural scenes and objects. *Journal of Vision, 3*, 75-85.
- 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.
- Vermaercke, B., & de Beeck, H. P. O. (2012). A multivariate approach reveals the behavioral templates underlying visual discrimination in rats. *Current Biology*, *22*(1), 50-55.
- 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.
- Wang, S., Tudusciuc, O., Mamelak, A. N., Ross, I. B., Adolphs, R., & Rutishauser, U. (2014). Neurons in the human amygdala selective for perceived emotion. *Proceedings of the National Academy of Sciences*, *111*(30), E3110-E3119.
- Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., et al. (2010). Does face inversion change spatial frequency tuning? *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 122–135.

## 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 *luminance* dimension (e.g., respond “mountain” when presented with a field scene); the red curves represent the proportion of errors on the *chrominance* 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). Adapted from Gosselin (2000).

*Figure 2.* Random-effects classification image illustrating how SF-time pixels correlate with accurate object recognition. Pixels enclosed by black lines are significant ( $p < .05$ , one-tailed).

*Figure 3.* This outlines the models of Bullier (2001b) and of Bar (2003). According to Bullier (2001b), 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.

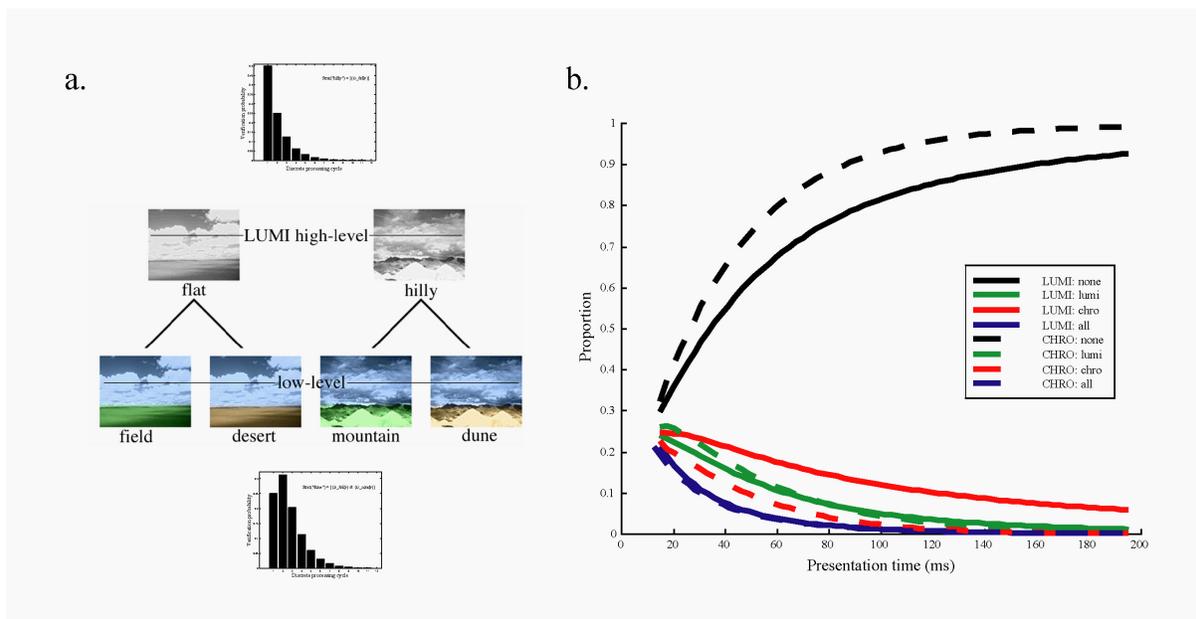


Figure 2.

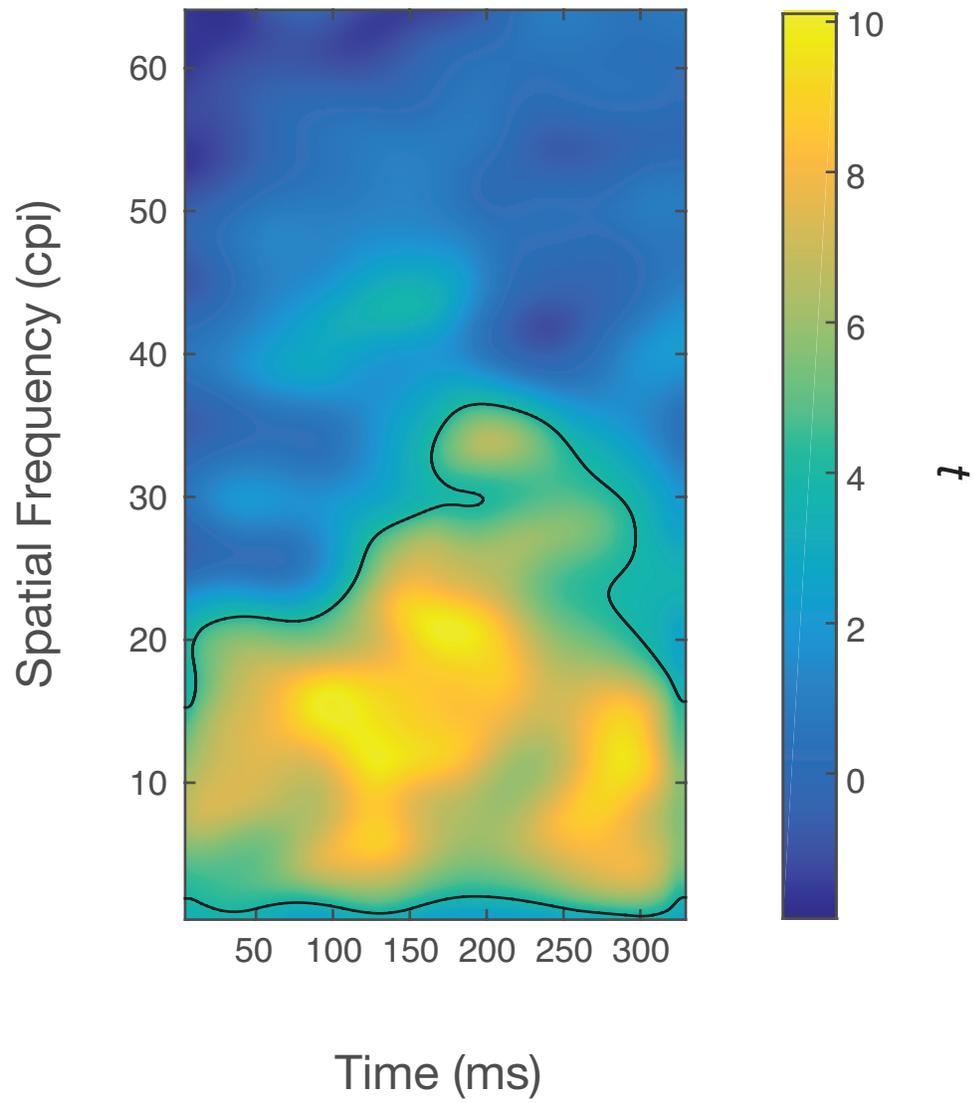


Figure 3.

