Indirect Influence of Taxonomic Knowledge on the Time Course of Scene Perception

Frédéric Gosselin
Université de Montréal
Philippe G. Schyns
University of Glasgow

Please send all correspondence to:
Frédéric Gosselin
Département de psychologie
Université de Montréal
C.P. 6128, succ. Centre-ville
Montréal QC
H3C 3J7 Canada
Phone: (514) 343-7550
Fax: (514) 343-2285
frederic.gosselin@umontreal.ca

Abstract

We synthesized four artificial scenes by combining two different luminance patterns (that we call *flat* and *hilly*) with two different chromatic patterns (labeled *grassy* and *sandy*). In a learning phase, all participants learned to categorize the four scenes at a general and at a specific level. At a general level, LUMI participants learned to separate the four scenes into "flat" and "hilly" on the basis of luminance cues; CHRO participants learned to separate the same scenes into "grassy" and "sandy" on the basis of chromatic cues. At a specific level of categorization, LUMI and CHRO participants learned to categorize the stimuli as either "field" (the combination of *flat* and *grassy*), "desert" (*flat* and *sandy*), "mountain" (*hilly* and *grassy*) or "dune" (*hilly* and *sandy*).

In a testing phase, we instructed participants to categorize the scenes at their most specific level (never at their general level). The error pattern of individual participants was successfully fitted to a three-parameter version of the *SLIP* model of categorization (Gosselin & Schyns, 2001). Observer groups assigned orthogonal weights to the luminance and chrominance dimensions (with the CHRO vs. LUMI group biased to the chromatic vs. luminance dimension) even though categorizations at the specific level (the task to resolve) was itself unbiased to one or the other dimension. This demonstrates that the abstract knowledge of category taxonomy can have indirect effects and modify the tuning of two primary dimensions of scene perception.

In the early days of vision research, it was commonly thought that abstract knowledge about the external world influenced its perception (e.g., Bruner & Goodnow, 1947; Klein, 1970). By "abstract knowledge" we mean knowledge derived from culture, values and language, rather than the sort of knowledge that is directly related to the visual world, and already embedded in the constraints and priors of mid- and high-level vision models (e.g., Marr, 1982; Knill & Richards, 1996; Gregory, 1970). Over the past decades, theories of vision driven by abstract knowledge have been slowly eliminated and replaced by "bottom-up" and encapsulated accounts (e.g., Pylyshyn, 1999). Nowadays, discoveries in the study of human knowledge rarely inform vision research, with the two fields drifting apart.

However, the idea that abstract knowledge can modify the tuning of early vision is appealing when dealing with visual stimuli as complex as scenes. To perform efficient scene categorizations, human perceivers in their natural environments must maximize their intake of meaningful visual information, but the limited bandwidth of visual pathways interferes with this goal (e.g., Broadbend, 1958; O'Regan, 1992). Despite limited bandwidth, human observers are known to categorize complex scene stimuli very quickly, often in a single glance (Antes et al., 1981; Biederman, 1981; Delorme et al., 2000; Hollingworth & Anderson, 2000; Potter, 1975; Schyns & Oliva, 1994; Thorpe et al., 1996), raising the issue of the determinants of efficient categorizations from sparse information.

Several direct determinants have been examined, including variations of luminance at different scales (e.g., Oliva & Schyns, 1997; Oliva & Torralba, 2001; Parker, Lishman & Hughes, 1992; Schyns & Oliva, 1994), but also variations of chromatic information (Goffaux et al., 2003; Oliva & Schyns, 2000). It was found that luminance and chromatic variations, when they are *directly* diagnostic of the category in question (e.g. the horizontal and vertical organization of buildings in *city*, or the yellow and blue contrast of sand and sea in *beach*) can enhance categorization performance. Sparse information, when it is directly diagnostic in the task, is one determinant of efficient scene categorization (Schyns, 1998).

However, there might be subtle, less direct effects of diagnosticity that could also determine speeded scene categorizations. It is now well established that scene categories are typically embedded, so that a beach is also "sandy thing", or a "flat thing" and such principle applies to all scene categories (Rosch, 1978). The critical issue is whether taxonomies of embedded categories (e.g. "desert, a sandy thing" or "desert, a flat thing") can *indirectly* influence the information extraction process. By indirect we mean that this influence will not arise from the direct diagnosticity of *desert* information when categorizing the input as "desert", but instead from the indirect knowledge that desert is a sandy thing, when categorizing the input as "desert". Briefly stated, we are addressing the issue of whether the diagnostic information of a higher-level category can influence the information extraction strategy required to perform a lower-level categorization.

Experiment

We synthesized four scene stimuli (field, hill, desert, dune) made from combinations of two luminance and two chromatic patterns (see Figure 1). All observers learned to categorize these scenes at a specific level as *field*, *hill*, *desert* and *dune*. Note that to do so, observers must combine the luminance and the chromatic information. The first observer subgroup also learned to label the scenes at an abstract level as "flat vs. hilly" on the basis of the luminance dimension (with *flat* = {field, desert} and *hilly* = {hill, dune}, see Figure 1). The second subgroup learned the abstract categorization "sandy vs. grassy" on the basis of the chromatic dimension (with *sandy* = {desert, dune} and *grassy* = {field, hill}, see Figure 1).

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. Suppose that observers are asked to categorize the briefly presented field picture *at the specific level*. In both groups, observers can make four responses: "field" suggests a correct perception of both dimensions (henceforth, called *none* "error"); "mountain" suggests a misperception of the flat luminance (henceforth, called *lumi* error); "desert" suggests a misperception of the green chrominance (henceforth, called *chro* error); and "dune" implies a misperception of both the flat luminance and the green chrominance of the field (henceforth, called *all* error)[1].

Indirect effects of diagnosticity imply that LUMI and CHRO observers would have different perceptions of identical scenes, when categorizing them at the same specific level. That is, observers placed in an identical condition of stimulation (e.g. seeing a field) and response (choosing between "field," "mountain," "desert" or "dune") would produce opposite patterns of categorization errors (i.e., respond more often "desert" than "mountain" in LUMI, but "mountain" than "desert" in CHRO), revealing a differential sensitivity to luminance and chrominance in the groups (see Figure 2). (A similar analysis applies to all four stimuli of the experiment.) This would happen if each group was tuned to chromatic and luminance information to maximize their categorization potential, even though the categorization at hand does not require such tuning (we come back to this point and its implications in the **Discussion**).

Method

Participants

Twenty-four University of Glasgow students with normal or corrected vision were paid to participate in the experiment.

Stimuli

[1] For the first three types of errors observers could also have responded correctly (e.g., respond "field" when presented with a field) or made an error on a single dimension (e.g., respond either "mountain" or "desert" when presented with a field) by chance alone; only when observers make errors on both dimensions (e.g., respond "dune" when presented a field) can we be sure that they misperceived both dimensions. Our model will take these into account.

We synthesized four distinct 450 x 350 pixels (spanning 7 x 5.4 deg of visual angle) stimuli–a field, a desert, a mountain, and a dune–with the Photoshop image processing software by combining two luminance patterns with two chromatic patterns. The luminance patterns were extracted from a field (called *flat* here) and a dune (*hilly* here) photographs from the Corel Draw Photo Database; they were normalized for size and horizon level. The chromatic patterns were composed of two colored rectangles corresponding roughly to the ground and the sky. The sky was the same blue, and the ground either green (called *grassy*) or yellow (called *sandy*). To eliminate the sharp boundary edge between the two colored rectangles, we low-passed the patterns. A mask was created by randomly assigning to each square of a 18 x 14 grid the content of the corresponding region of one of the four scenes.

Procedure

The experiment ran on a Macintosh Power PC using a program written with the Psychophysics Toolbox for Matlab (Brainard, 1997; Pelli, 1997). Stimuli were presented on a high-resolution calibrated monitor. Two observer groups (called LUMI and CHRO) learned the name of the four stimuli at the specific level of a taxonomy: "field," "mountain," "desert" and "dune." LUMI observers also learned to categorize the stimuli at a general taxonomic level into "flat" vs. "hilly," on the basis of luminance cues, whereas CHRO observers learn to categorize the stimuli at a general level into "grassy" vs. "sandy," using chromatic cues (see Figure 1).

A learning block was completed when participants had named consecutively–and without mistake–all scenes at the high and at the low levels of categorization (4 scenes * 2 levels of abstraction = 8 trials minimum), with LUMI and CHRO differing only in their high-level categorizations. Trial order was randomized within each learning block. A trial began with the display of the word "high" or "low," instructing observers of the level at which the subsequent scene had to be named. Observers then pressed a key to display the scene to categorize (presented on the screen for 1 s) immediately followed by a 450 ms mask. Observers indicated their categorization using one of six response-keys (two for the high level, four for the low level) before moving on to the next trial. Corrective feedback was provided.

When participants reached criterion (all observers reached criterion after exposition to only four general- as well as four specific- levels trials, the minimum number – this is important and we will get back to this in the **Discussion**), they were transferred to a testing phase including trials differing in three ways from those described above: First, we randomly varied presentation time from trial to trial (either 15, 45, 75, 105, 135, 165, or 195 ms) to derive a range of performance. Secondly, we only tested the low level of categorization, never the high-level one. That is, each one of the 700 test trials (4 scenes x 7 presentation times x 25 repetitions presented in a random order) started with the word "low". Thirdly, no corrective feedback was given.

Results

There are four types of error (*none, lumi, chro,* and *all*) and seven presentation times, for a total of 28 data points per observer. Because the probabilities of the four error conditions add up to 1, however, we really only have 21 independent data points – estimated by a total of 700 trials – per observer. We thus fitted the 21 independent data points of each individual observer using an adapted version of *SLIP* (Gosselin & Schyns, 2001). With an adequate model, this approach enables an optimal use of the information contained in the observers' data to infer perceptual weights. The model is fully described in the Appendix. Here, we only reproduce the set of equations providing the probabilities of the four types of error as a function of *t*, the presentation time in clock cycles: $P(all) = .25k^{tw_{lumi}}k^{tw_{chro}}$, the proportion of errors on *all* dimensions (e.g., respond "dune" when presented with a field); $P(lumi) = .5k^{tw_{chro}}\left(1-.5k^{tw_{lumi}}\right)$, the proportion of errors on the *lumi*nance dimension (e.g., respond "mountain" when presented with a field scene); $P(chro) = .5k^{tw_{lumi}}\left(1-.5k^{tw_{chro}}\right)$, the proportion of errors on the *chro*minance dimension (e.g., respond "desert" when presented with a field scene); and $P(none) = \left(1-.5k^{tw_{lumi}}\right)\left(1-.5k^{tw_{chro}}\right)$, the proportion of errors on *none* of the perceptual dimensions (e.g., respond "field" when presented with a field scene). The constant *k* was arbitrarely set to .5 (this constant has very little impact on the fitting results). These four equations have only three free parameters: the first one is w_{lumi}, the perceptual weight given to the luminance dimension (w_{chro}, the perceptual weight given to the chrominance dimension is equal to 1 - w_{lumi}, with $0 < w_{lumi} < 1$), and the last two are linear scaling parameters (i.e., $RT = at + b$). We used an implementation of the Nelder-Mead simplex algorithm for Matlab from the *MM5 Toolbox* (Hanselman & Littlefield, 1998) to maximize the goodness of fit. We obtained remarkably large R^2s (ranging from .92 to .99, with an average of .98).

CHRO observers weighted the chrominance dimension more heavily (0.599) than to the luminance dimension (0.401), whereas LUMI observers weighted the luminance dimension more heavily (0.599) than the luminance dimension (0.401). Out of the 24 participants, 19 had weight differences (i.e., $w_{lumi} - w_{chro}$) with the expected sign ($p < .01$).

Figure 3 shows the 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). At 15 ms

exposure, performance is near chance; it quickly rises above chance for longer exposures. Remarkably, the two *lumi* lines as well as the two *all* lines overlap almost perfectly throughout presentation times. The two *chro* lines as well as the two *none* lines diverge around 20 ms and remain apart until 195 ms.

General Discussion

We synthesized four artificial scenes by combining two different luminance patterns (that we call *flat* and *hilly*) with two different chromatic patterns (labeled *grassy* and *sandy*). The LUMI observers learned to separate the four scenes into "flat" and "hilly" on the basis of luminance cues at the most general level of categorization, whereas CHRO observers learned to separate them into "grassy" and "sandy" on the basis of chromatic cues. Both the LUMI and the CHRO participants learned to categorize the stimuli as either "field" (the combination of *flat* and *grassy*), "desert" (*flat* and *sandy*), "mountain" (*hilly* and *grassy*) or "dune" (*hilly* and *sandy*) at the most specific level of categorization. During the actual experiment, they were only asked to categorize the scenes at the most specific level of categorization.

As predicted, CHRO observers weighted the chrominance dimension more heavily than to the luminance dimension, whereas LUMI observers weighted the luminance dimension more heavily than the luminance dimension. This would happen if each group was tuned to chromatic and luminance information to maximize their categorization potential (e.g., Rosch, 1978). After a test of only 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.

This effects illustrates the indirect, top-down influence of a taxonomy, as discussed earlier. Both groups performed exactly the same experiment, except for the dimension structuring the abstract categories in the groups. Still, could some sort of "bottom-up" learning – e.g., perceptual learning, Falhe & Poggio, 2002) – of the chrominance (vs. luminance) patterns have taken place when observers categorized the stimuli at the higher level? It seems unlikely. None of the participants was submitted to more than four such high-level categorizations. The effect of four trials seems rather negligeable compared to that of the 700 low-level categorizations that both groups had to complete. It seems even more negligible compared to the hundreds of thousands of times that natural scenes are categorized as "flat", "hilly", "sandy", or "grassy" by a typical human observer during her or his early adulthood.

This raises the question of the locus of the change induced by the abstract taxonomic knowledge. The simpliest hypothesis involves an alteration in the speed (or efficiency) of information processing along the two visual dimensions of luminance and of chrominance (e.g., Livingstone & Hubel, 1988). This would account for the earliness of the effect (demonstrated as early as 20 ms after stimulus onset). There is a weakness with this hypothesis. Remember that both *lumi* (e.g., respond "mountain" when presented with a field)

lines in Figure 3 overlapped almost perfectly indicating that observers from both groups behaved as if they processed luminance equally fast. In fact, the difference in weighing of the luminance and chromatic information in the groups arises essentially from differences in processing the chrominatic dimension—observe the divergence of the *chro* (e.g., respond "desert" when presented with a field) lines in Figure 3. Observers in the CHRO group behaved as if they were more efficient at processing chromatic information than those of the LUMI group. It is logically possible that the low-level processing speed of chrominatic – but not that of luminance – information is adjustable. But why would luminance and chrominance be qualitatively different?

A better hypothesis is that all observers process chromatic information faster than luminance. In such a case, the maximum processing speed of luminance would act as a bottleneck. The optimal strategy would be to modify the processing speed of chrominance according to task demands, and to always process luminance as fast as possible. The CHRO observers would process both luminance and chrominance at full throttle. And those of the LUMI group would process luminance at full throttle and chrominance a little slower than luminance. There is evidence for the crux of this hypothesis in the literature. Moutoussis & Zeki (1997a, 1997b) have shown that color, form, and motion are perceived separately, and in this order, from fastest to slowest (see also Holcombe & Cavanagh, 2001). In any case, we believe that our results demonstrate that abstract taxonomic knowledge can modify the time course of low-level scene perception.

Concluding remarks

We argued earlier that knowledge and perception should be reunited, minimally because the former offers an organization of visual information to the observer. This study showed that a pervasive principle of organization of abstract knowledge (a taxonomy of categories) could indirectly and selectively modify the perception of an identical scene input along two of its primary dimensions (luminance and chrominance). This rehabilitation of the role of abstract knowledge in scene perception is an existence proof. It paves the way for new research on the interactions between the knowledge-driven expectations of the observers and the natural constraints of the visual world.

Appendix

Here, we adapt the Gosselin & Schyns (2001) *SLIP* to model the current experiment. *SLIP* is an ideal categorizer that applies "optimal" testing strategies to determine the category membership of objects (for a similar approach see Feldman, 2000). A strategy comprises sets of noisy detectors (e.g. Strat(*desert*) = [{*is_flat*} & {*is_sandy*}], is the *SLIP* strategy for desert which comprises two sets of detectors of one element each; Strat(*flat*) = [{*is_flat*}] is the strategy which comprises one set of detector of one element). *SLIP* launches all these detectors in parallel. 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" one successful luminance detector suffices; and to verify that that a scene is "grassy" one successful chrominance 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 clock cycles (*t*). We assume that response time (*RT*) is a linear function of the number of clock cycles (i.e., $RT = a\ t + b$, with *a* and *b*, two free parameters). The redundancy of the set of luminance detectors (w_{lumi}; note that we defined the redundancy of the set of chrominance detector as $w_{chro} = 1 - w_{lumi}$) is the most important and the last free parameter in our model.

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" 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 clock 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 clock cycle *t* or before is:

$$\prod_{j=1}^{n}\left(1-\phi_j^t\right) \qquad \text{(Equation 1)},$$

with $\phi_j = k^{w_j}$. The constant *k* is the probability that one detector is successful during one clock cycle. It was arbitrarily set to .5 for the bestfits reported in this article. The other constant, w_j, is an index of the redundancy of set *j*; it is the weight given to dimension *j*. The primary aim of our curve fitting efforts is to find the value of w_{lumi} (and, by the same token, w_{chro} defined as $1 - w_{lumi}$) for each participant.

For the purpose of this article, we only need three particular instances of Equation 1: $\varphi_{lumi} = \left(1-\phi_{lumi}^t\right)$, the cumulative probability that at least one luminance detector will be successful at *t*; $\varphi_{chro} = \left(1-\phi_{chro}^t\right)$, the cumulative probability that at least one chrominance detector will be successful at *t*; and $\varphi_{none} = \left(1-\phi_{lumi}^t\right)\left(1-\phi_{chro}^t\right)$, the cumulative probability that at least one luminance detector and one chrominance detector will be successful at *t*.

If our observers were not required to guess when they don't perceive, φ_{none}, $\varphi_{lumi}\neg\varphi_{chro}$ ($\neg\varphi_x = 1-\varphi_x$), $\neg\varphi_{lumi}\varphi_{chro}$, and $\neg\varphi_{lumi}\neg\varphi_{chro}$ could be used directly to model their error patterns that is, the cumulative probability that at least one luminance detector and one chrominance detector is successful at *t* (*P(none)*), the cumulative probability that at least one luminance detector and no chrominance detector is successful at *t* (*P(chro)*), the cumulative probability that no luminance detector and one chrominance detector is successful at *t* (*P(lumi)*), and the cumulative probability that no detector at all is successful at *t* (*P(all)*). Observers, however, were required to provide an answer, hence to guess. Next, we correct our model for guessing.

Observers can make errors on *all* perceptual dimensions (e.g., respond "dune" when presented with a field scene) only if they guessed incorrectly both dimensions and all their dimension detectors were unsuccessful. The probability of the former event is .25 because there are four low-level categories; the probability of the latter is given by $\neg\varphi_{lumi}\neg\varphi_{chro}$. Therefore, *P(all)* = $.25\neg\varphi_{lumi}\neg\varphi_{chro}$. Putting everything together and simplifying: *P(all)* = $.25k^{tw_{lumi}}k^{tw_{chro}}$.

There are two ways for observers to make an error on the luminance – *lumi* – dimension: either all their detectors are unsuccessful ($\neg\varphi_{lumi}\neg\varphi_{chro}$) and the chrominance – but not the luminance – is guessed correctly (.5 * .5 = .25), or at least one chrominance detector is successful (φ_{lumi}) and luminance is neither tested nor guessed correctly ($.5\neg\varphi_{chro}$). Thus, *P(lumi)* = $.25\neg\varphi_{lumi}\neg\varphi_{chro} + .5\varphi_{lumi}\neg\varphi_{chro}$. Similarly, we derive *P(chro)* = $.25\neg\varphi_{lumi}\neg\varphi_{chro} + .5\neg\varphi_{lumi}\varphi_{chro}$. Putting everything together and simplifying: *P(lumi)* = $.5k^{tw_{chro}}\left(1-.5k^{tw_{lumi}}\right)$ and *P(chro)* = $.5k^{tw_{lumi}}\left(1-.5k^{tw_{chro}}\right)$.

Finally, four roads can lead an observer toward misperception of *none* of the dimensions: (1) at least one luminance detector and one chrominance detector are successful (φ_{none}), (2) at least one luminance detector is successful and – although no chrominance detector is successful – chrominance is correctly guessed ($.5\varphi_{lumi}\neg\varphi_{chro}$), (3) at least one chrominance detector is successful and – although no luminance detector is successful – luminance is correctly guessed ($.5\neg\varphi_{lumi}\varphi_{chro}$), and (4) all detectors are unsuccessful but – nonetheless –both luminance and chrominance values are correctly guessed ($.25\neg\varphi_{lumi}\neg\varphi_{chro}$). Summing up everything, we end up with: *P(none)* = $\varphi_{none} + .5\varphi_{lumi}\neg\varphi_{chro} + .5\neg\varphi_{lumi}\varphi_{chro} + .25\neg\varphi_{lumi}\neg\varphi_{chro}$. Putting everything together and simplifying: *P(none)* = $\left(1-.5k^{tw_{lumi}}\right)\left(1-.5k^{tw_{chro}}\right)$.

References

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

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

Antes, J. R., J. G. Penland, et al. (1981). Processing global information in briefly presented pictures. *Psychological Research, **43***, 277-92.

Biederman, I. (1981). On the semantics of a glance at a scene. Perceptual Organization. M. Kubovy and J. R. Pomerantz. Hillsdale, NJ, Erlbaum**:** 213-233.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433-436.

Broadbend, D. E. (1958). *Perception and communication.* London: Pergamon Press.

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

Bruner, J. S. & Goodman, C. C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology, 42*, 33-44.

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

Falhe, M. & Poggio, T. (Eds.) (2002). *Perceptual learning.* Massachussetts: MIT Press.

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

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

Gregory, R. (1970). *The intelligent eye.* New York: McGraw-Hill.

Hanselman, D. & Littlefield, B. (1998). *Mastering Matlab 5.* New Jersey: Prentice Hall.

Holcombe, A. O. & Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nature Neuroscience, 4*, 127-128.

Klein, G. S. (1970). *Perception, motives, and personality.* New York: Alfred A. Knopf.

Knill, D. C. & Richards, W. (1996). *Perception as Bayesian inference.* Cambridge: Cambridge University Press.

Livingstone, M. & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and psychology. *Science, 240*, 740-749.

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

Moutoussis, K. & Zeki, S. (1997a). Functional segregation and temporal hierarchy of the visual perceptive systems. *Proceedings of the Royal Society of London B, 264*, 1407-1414.

Moutoussis, K. & Zeki, S. (1997b). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society of London B, 264*, 393-399.

Nao-Raz, G., Tarr, M. J. & Kersten, D. (2003). Is color an intrinsic property of object representation? *Perception, 32*, 667-680.

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

Oliva, A. & Schyns, P. G. (2000). Colored diagnostic blobs mediate scene recognition. *Cognitive Psychology, 41*, 176-210.

O'Regan, J. K. (1992). Solving the "real" problem of perception: the world as an outside memory. *Canadian Journal of Psychology, 46*, 461-488.

Parker, D. M., J. R. Lishman, et al. (1992). Temporal integration of spatially filtered visual images. *Perception, 21*, 147-60.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437-442.

Potter, M. C. (1975). Meaning in visual search. *Science, 187*, 965-966.

Pylyshyn, Z. (1999). Is vision continuous with cognition? – The case for Cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22*, 341-423.

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.

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

Schyns, P. G. (1998). Diagnostic recognition: Task constraints, object information and their interactions. *Cognition*, *67*, 147-179.

Schyns, P. G. and A. Oliva (1994). From blobs to boundary edges: evidence for time and spatial scale dependent scene recognition. *Psychological Science, 5*, 195-200

Tanaka, J., Weiskopf, D. & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences, 5*, 211-215.

Thorpe, S., D. Fize, et al. (1996). Speed of processing in the human visual system. *Nature, 381*, 520–522.

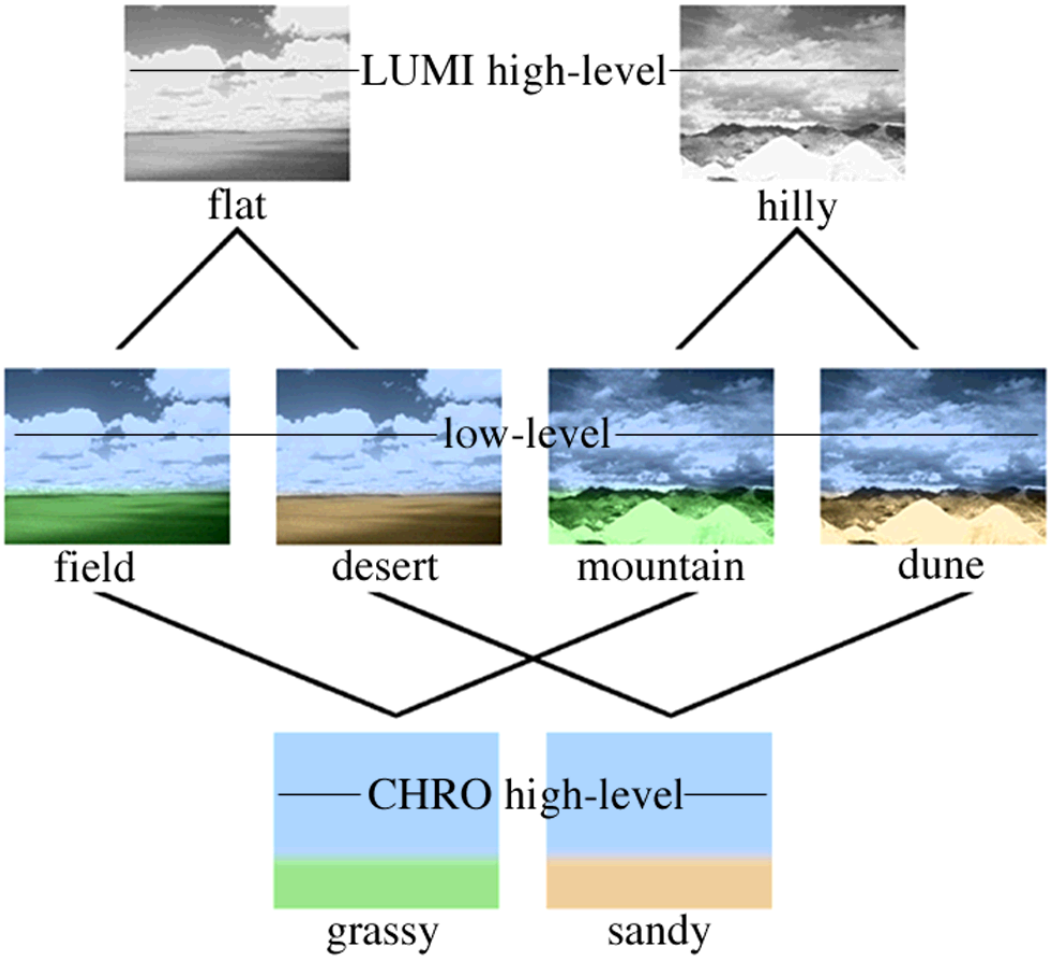

Figure 1. The four scenes used in this experiment and the corresponding low-level category names learned by all participants ("field", "mountain", "desert", and "dune"), sandwiched by the two high-level categorizations ("flat" and "hilly") LUMI observers learned, and those ("grassy" and "sandy") CHRO observers learned. Note that each scene is the conjunction of a luminance and a chrominance pattern.

Figure 2. An illustration of the possible categorization errors when an observer is presented with a field (i.e., *flat* and *grassy*) scene: "field" suggests that *none* of the perceptual dimensions were misperceived (the correct response could also have been reached by guessing – see *Results* and *Appendix* for a discussion of the guessing issue); "mountain" suggests that *lumi*nance was misperceived but not chrominance; "desert" suggests that *chro*minance was misperceived but not luminance; and "dune" implies that *all* perceptual dimensions were misperceived.

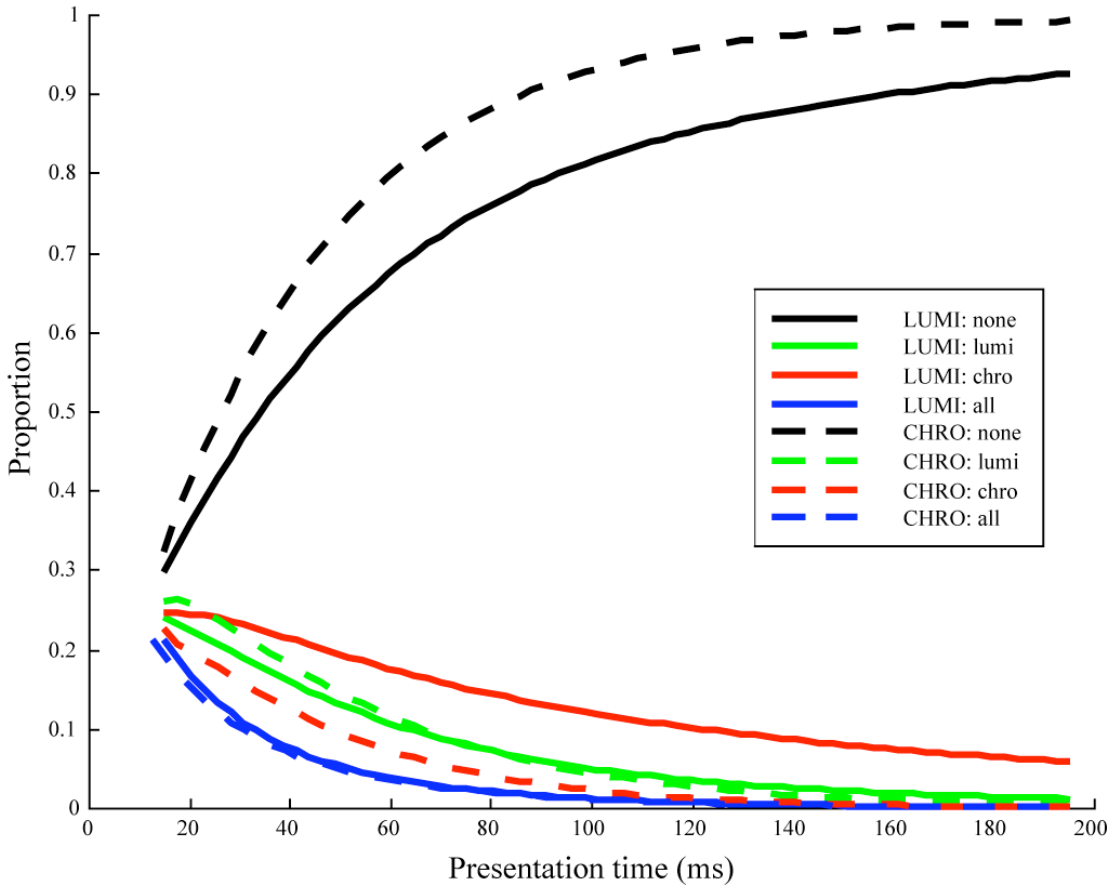

Figure 3. Proportion of the four error types in function of presentation time for each observer group. The solid and dashed lines are, respectively, the average bestfits of the *SLIP* model to the LUMI and to the CHRO participants individual data points. The black curves represent the proportion of errors 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).