Perception, 2002, volume 31, pages 683 – 691

DOI:10.1068/p3276

Understanding Dali's *Slave Market with the Disappearing Bust of Voltaire*: A case study in the scale information driving perception

Lizann Bonnar¶, Frédéric Gosselin§, Philippe G Schyns¶
Department of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, UK;
§ Present address: Département de Psychologie, Université de Montréal, Montréal, Québec H3C 3J7, Canada; e-mail: lizann@psy.gla.ac.uk, philippe@psy.gla.ac.uk
Received 31 August 2001, in revised form 8 March 2002

Abstract. A generic problem in vision is to know which information drives the perception of a stimulus. We address this problem in a case study that involves the perceptual reversal of an ambiguous image (here, Dali's painting the *Slave Market with the Disappearing Bust of Voltaire* 1940). In experiment 1, we use 'bubbles' (Gosselin and Schyns, 2001 *Vision Research* **41** 2261 – 2271) to disambiguate the image and to determine the specific visual information that drives each possible perception (here, the nuns versus the bust of Voltaire). Experiment 2 validates that this information does determine the selective perception of the ambiguous image. We adapted the spatial-frequency channels of observers selectively to the information that mediates one of the two perceptions, to induce the opposite perception of the ambiguous image in a transfer phase. Together, the results suggest a method of revealing the visual information that drives perception.

1 Introduction

The perceptual reversal of ambiguous images has been a source of fascination for psychologists and artists alike, albeit for rather different reasons. For some artists, the allure in introducing ambiguity is to create in the observer an experience that is, explicitly, purely subjective and qualitative. It is a way of emphasising the constructive nature of perception, the observer's share. For the psychologist, on the other hand, image ambiguity serves as a tool to probe the dynamics of the visual and cognitive system: the retina receives a single image comprising multiple interpretations, yet the visual and cognitive system is constrained so that only one percept is available at a time.

Classical examples of bi-stable ambiguous figures include some based on figure – ground, eg Rubin's faces – vase figure; some perspective-based, such as the Schroeder staircase and the Necker cube; and others, object ambiguities, such as the duck – rabbit figure. The use of such ambiguities is evident in the work of many artists; these 'double-images' were in fact a favorite tool of the surrealist painter, Salvador Dali (Descharnes 1972). One such example of Dali's paintings is *Slave Market with the Disappearing Bust of Voltaire* (1940) in which the heads of two nuns within a busy scene also constitute the eyes of the Bust of Voltaire. On viewing this painting, perception switches between one interpretation and the other.

A general problem in vision that also applies to these bi-stable ambiguous images is to know which information drives perception. Ambiguous images, such as Dali's painting, are perfect stimuli for investigating the information in a stimulus that underlies its perception; the bottom – up information is identical. However, up until recently, no generic technique was available to simplify a stimulus to the essential information driving its perception. Here, we adapt 'bubbles' (Gosselin and Schyns 2001), a technique developed to isolate recognition information, to visualise and validate the information driving the perception of each image interpretation in Dali's painting.

¶ Authors to whom all correspondence and requests for reprints should be addressed.

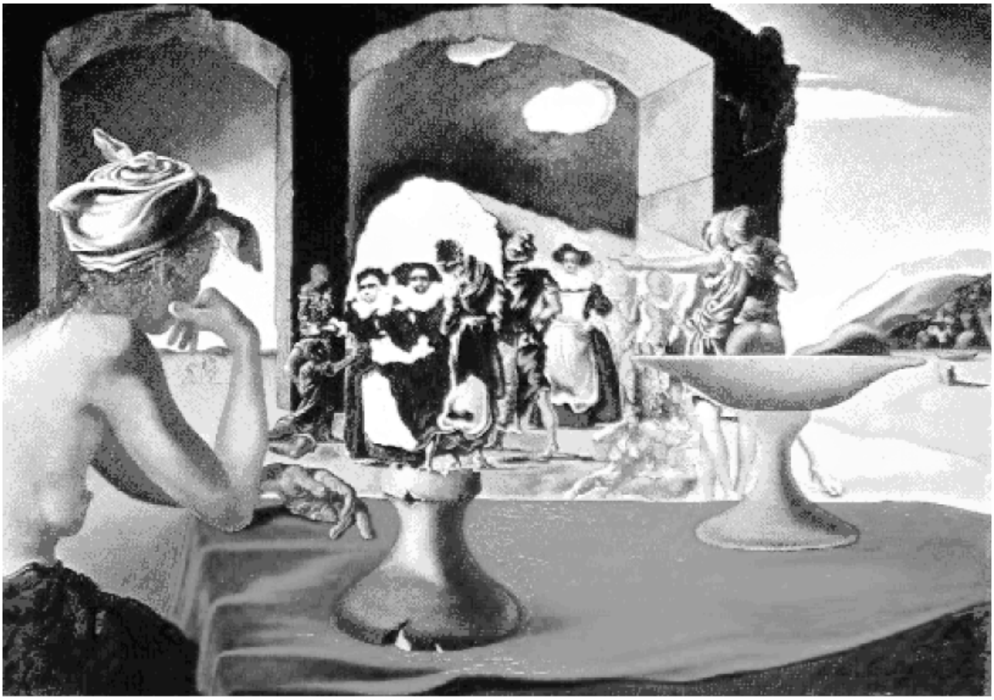

Figure 1. A grey-scale version of Salvador Dali's *Slave Market with the Disappearing Bust of Voltaire* (1940).

In experiment 1, we sought to reveal the information underlying each stable percept of the bi-stable ambiguous figure contained within Dali's painting. We found that the nature of this information is grounded in different spatial filters 'looking at' the image. In experiment 2, we validated these results using an established psychophysical technique known as frequency-specific adaptation (eg Blakemore and Campbell 1969; De Valois and De Valois 1990).

2 Experiment 1

To determine the use of information specific to each percept, we applied the Gosselin and Schyns (2001) 'bubbles' technique. Our method randomly searches an image-generation space (here, we chose the 2-D image plane and spatial scales) to present sparse versions of the image as stimuli.

Observers respond according to their percept of the sparse stimuli and we retain the samples of information that lead to the "nuns" response and the "Voltaire" response. From this information, we can establish and demonstrate which visual information selectively drives the percept of each image interpretation.

2.1 *Method*

2.1.1 *Participants.* Participants were ten University of Glasgow students with normal, or corrected-to-normal, vision, who were paid for their participation.

2.1.2 *Stimuli.* The experiment ran on a Macintosh G4 with a program written with the Psychophysics Toolbox for Matlab (Brainard 1997; Pelli 1997) and the Matlab Pyramid Toolbox (Simoncelli 1997). To determine the information driving each percept, we used 'bubbles' to generate sparse stimuli that randomly sampled an image-generation space comprising the 2-D image plane and a range of six spatial-frequency bandwidths (see figure 2 for a depiction of how the stimuli were generated). These independent bands of spatial frequencies of one octave each had cut-offs at 128, 64, 32, 16, 8, and

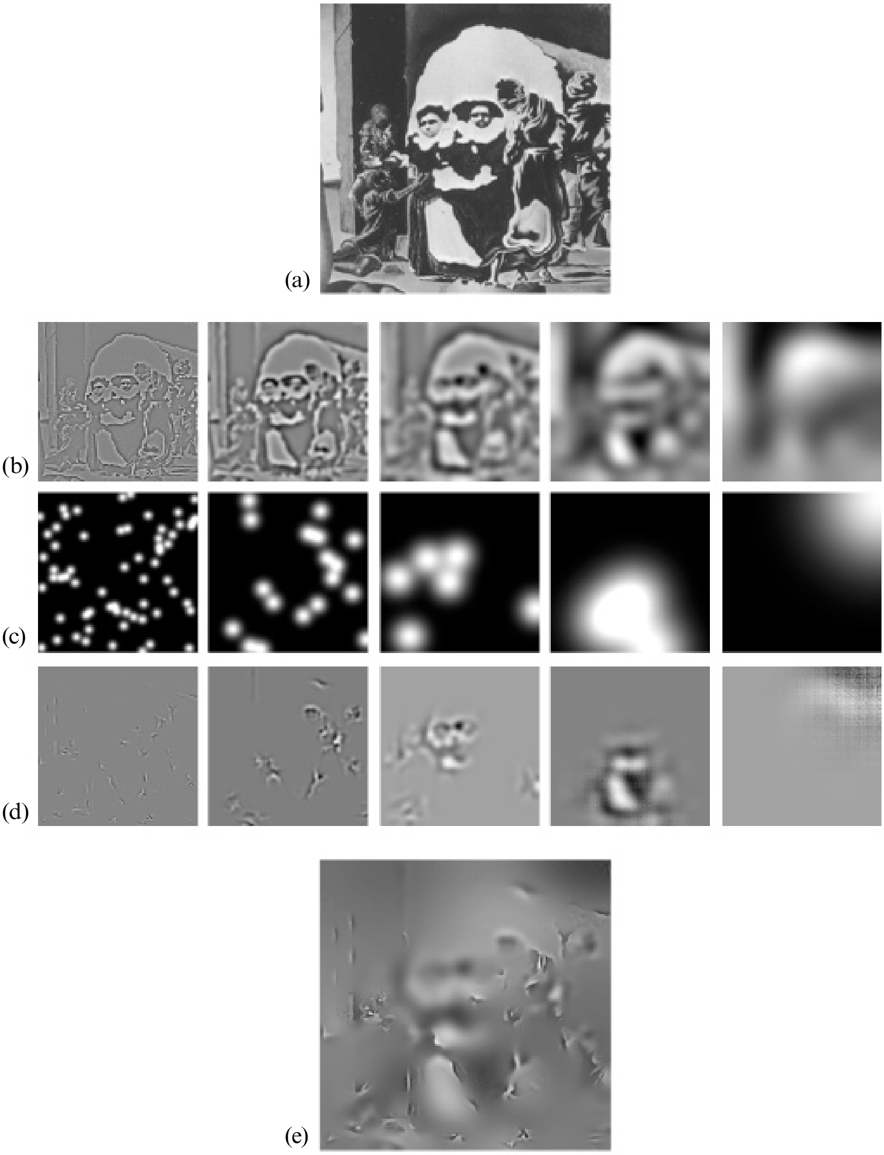

Figure 2. This figure illustrates the application of 'bubbles' to the 3-D image-generation space, comprising the 2-D ambiguous image and the decomposition into different spatial frequencies. Pictures in (b) represent five different scales of (a); (c) illustrates the bubbles applied to each scale and (d) shows the revealed information of (b) by the bubbles of (c). By reconstructing the information in (d) we obtain (e) an example of stimuli that subjects actually saw. Here subjects typically report seeing Voltaire.

4 cycles per image, from fine to coarse scales (22.378, 11.189, 5.594, 2.797, 1.399, and 0.699 cycles deg^{-1}). The coarsest band was not searched for perceptual information; instead it served only as a constant background. The image was presented centrally on the screen and the background luminance was 14 cd m^{-2}.

The image that was represented at each bandwidth was partially revealed by a mid-grey mask that was perforated by a number of randomly located Gaussian windows called 'bubbles' (see figure 2). We normalised to three the number of cycles that any bubble could reveal and adjusted the size of the bubble for each frequency band accordingly (the standard deviations of these bubbles were 0.13, 0.27, 0.54, 1.08, and 2.15 cycles deg^{-1}, from fine to coarse scales). Since the size of the bubbles increases from fine to coarse scales, the number of bubbles at each scale was adjusted to maintain constant, on average, the total area of the image revealed.

2.1.3 *Procedure.* We cropped the ambiguous portion of the grey-scale version of Dali's *Slave Market with Disappearing Bust of Voltaire* so that it comprised the bust of Voltaire and the two nuns (see the top picture of figure 2), and subtended 5.72 deg $\times$ 5.72 deg on the screen (for an image of 256×256 pixels). On any trial, participants were instructed to indicate by appropriate key presses which image they could perceive, the nuns or Voltaire. In the event that there was simply insufficient information to perceive either percept, subjects were instructed to press the "don't know" key. We introduced the "don't know" response as a tool to adjust on-line, on a trial-per-trial basis, the total number of bubbles sampling the image-generation space so that the number of "don't know" responses did not exceed 25%. It was emphasised that this response was a last resort that should not be used when both image interpretations were available. In this case, participants had to choose the strongest percept. The experiment comprised 500 trials, and in each trial a sparse image computed as described appeared on the screen. A chin-rest maintained viewing distance at 100 cm.

2.2 *Results and discussion*

Across trials, we keep track of the locations of the bubbles that lead to each image interpretation, the nuns and Voltaire. To this end, we created a different NunsPlane per scale [for each one of the five scales, henceforth NunsPlane(scale)], and a different VoltairePlane per scale [henceforth VoltairePlane(scale)]. Whenever participants perceived the nuns (versus Voltaire), we literally added the masks of bubbles to the NunsPlane(scale)

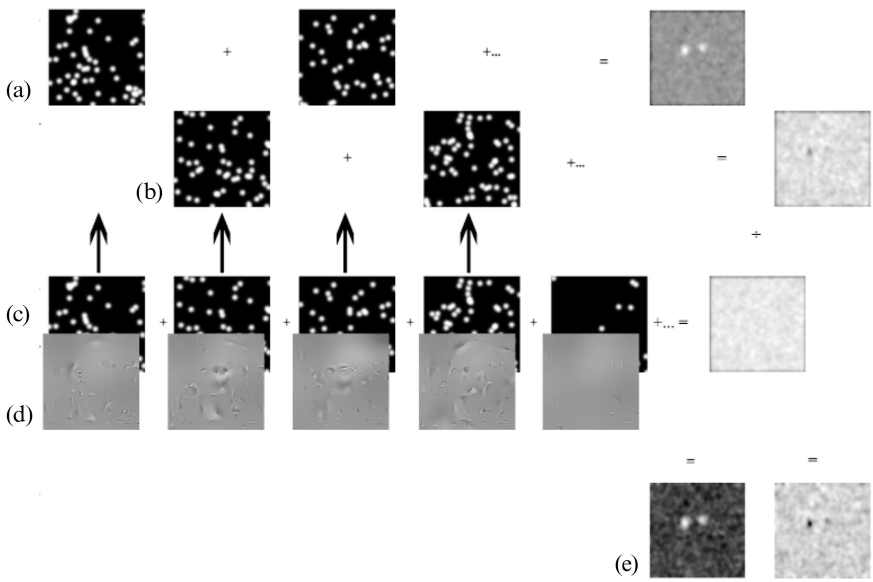

Figure 3. Computing the significance of each region in driving each percept at the finest scale. Identical computations are performed at the four other bandwidths. The pictures in (a) depict the addition of the bubbles that led to a "nuns" response to the NunsPlane, the top rightmost picture. The pictures in (b) illustrate the addition of the bubbles that led to a "Voltaire" response; the rightmost picture is the outcome of this addition, VoltairePlane. In (c) all bubbles (those leading to a "nuns", a "Voltaire", and a "don't know" response) are added to form TotalPlane (the rightmost grey-scale picture). In (d) are shown examples of experimental stimuli as revealed by the bubbles of (c). NunsProportionPlane and VoltaireProportionPlane in (e) from left to right, respectively, is the division of NunsPlane by TotalPlane and VoltairePlane by TotalPlane. (Note the whiter area in the NunsProportionPlane corresponding to the heads of the two nuns.)

[versus VoltairePlane(scale)], to keep track of the scale-specific bubbles of information leading to these perceptions (see figure 3). We also added the bubbles leading to "don't know" responses to the DontKnowPlane(scale). Together, the NunsPlane(scale), VoltairePlane(scale), and DontKnowPlane(scale) encode how frequently each subspace of the 3-D-image-generation space was interpreted as nuns, Voltaire, or "don't know".

To perform a statistical analysis, we first divide the frequencies in each plane by the total number of presentations to derive a proportion plane per scale. We then constructed a confidence interval ($p < 0.01$) around the mean for each of these proportions at each scale, NunsProportionPlane(scale) (M = 0.341, SD = 0.092) and for Voltaire-ProportionPlane(scale) (M = 0.425, SD = 0.078). We disregarded all of the areas below this confidence interval, as they were considered unimportant in driving each percept. Figure 4 depicts the selectively attended information driving each percept, that is those regions above the confidence interval.

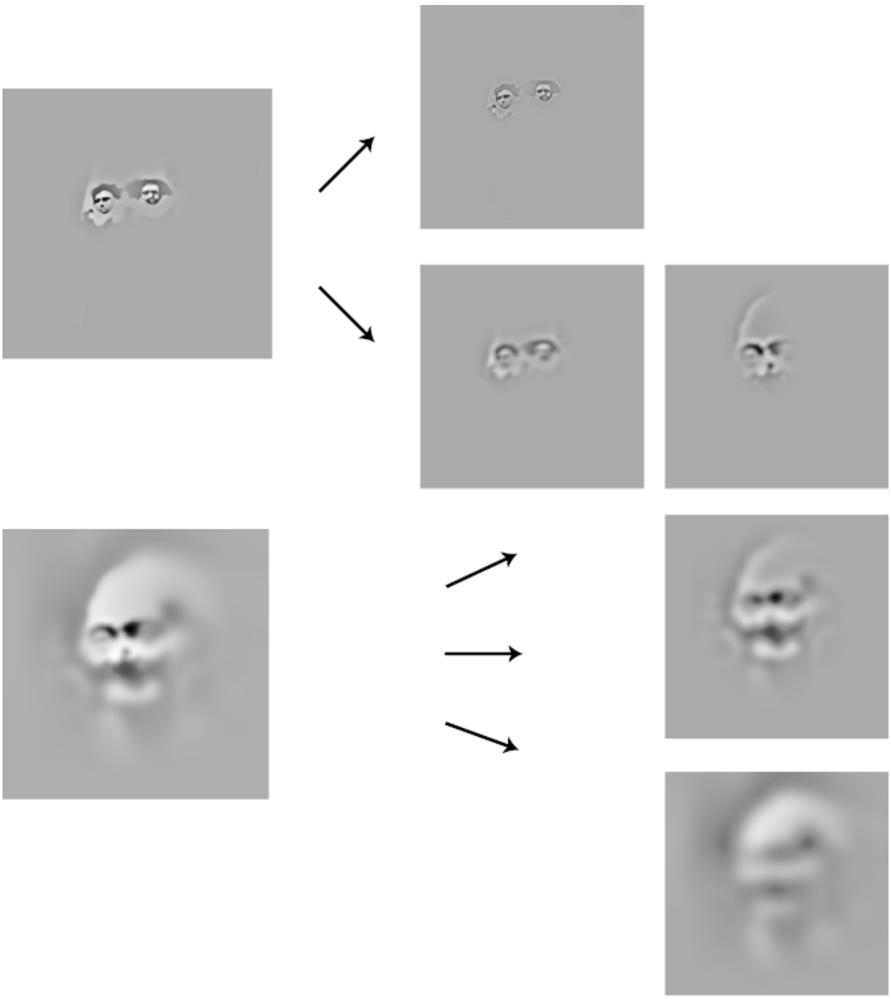

Figure 4. The potent information that drives the perception of the nuns and Voltaire is depicted in the leftmost column. The top leftmost image depicts the potent information driving the nuns percept. The two rightmost columns are decompositions of this information into the four scales found to be diagnostic, from finest at the top to the fourth scale at bottom. The second column demonstrates the information relevant to the nuns, and the third column, the information relevant to Voltaire.

Figure 4 presents the information driving the stable percepts of Dali's ambiguous painting (see figure 1). The top (versus bottom) pictures depict the information driving the nuns (versus Voltaire) percept. Note that the nuns occupy the finest and second-to-finest scales with no relevant information at the remaining three coarser scales and comprise the heads of the two nuns. In contrast, Voltaire dominates the second, third, and fourth scales, with the coarser scales encompassing the entire bust. Clearly, the information driving the two image interpretations differs across the 2-D image plane and across different bandwidths of the spatial spectrum, with an overlap only at the second scale.

The differential scale contribution in driving the alternative percepts in Dali's painting suggests that the alternation between each image interpretation may be explained in terms of a switch between the spatial filters 'looking at' the image (Oliva and Schyns 1997; Schyns and Oliva 1999).

Other ambiguities in art have also been explained by spatial-frequency accounts: for example, recently Livingstone (2000) suggested that the ambiguous expression in da Vinci's portrait of the Mona Lisa appears differently according to different foveal eccentricities, her smile being more apparent in the low spatial frequencies. Hayes and Ross (1995) also discuss the role of fine and coarse spatial scales in the efficiency of line drawings in representing complex scenes.

3 Experiment 2

In experiment 1, 'bubbles' isolated the spatial-scale information that selectively drives the perception of an ambiguous figure. However, this method is indirect, relying on the participants' categorisations of sparse samples of information that could induce strategies atypical of perception. To strengthen our results, we must establish a direct link between the isolated scale information and classical mechanisms of perception.

To this end, in experiment 2 we seek to ground the scale-specific perceptions in early vision, using an established psychophysical technique known as frequency-specific adaptation (eg Blakemore and Campbell 1969; De Valois and De Valois 1990). The rationale of frequency-specific adaptation is that an adaptation to pattern X changes the appearance or sensitivity to X, but not the appearance or sensitivity to pattern Y, thus indicating that the underlying structures simultaneously process independent aspects of the patterns. For example, Blakemore and Campbell (1969) showed that observers exposed to a sine-wave pattern oscillating at, eg, 5 cycles $\deg^{-1}$ exhibited a reduction in their ability to perceive contrast at this particular frequency. That is, adaptation to a spatial frequency selectively impaired sensitivity to this particular frequency, as if only this channel was affected, but not the others (see also Pantle and Sekuler 1968).

If perception results from a switch between distinct spatial channels, then we can apply this reasoning to the selective perception of the ambiguous Dali painting. In experiment 2 we use frequency-specific adaptation to selectively turn off the spatial filters underlying one percept (eg Voltaire) in order to disambiguate the picture (eg induce a stable perception of only the nuns). It is important to stress that our goal is not to adapt the visual system to a specific percept (ie nuns or Voltaire), but rather to habituate the mechanisms of early vision (ie spatial-frequency channels) that mediate this percept. In a series of experiments, Webster and Miyahara (1997) adapted observers to natural images and measured the effect of the adaptation on the contrast-sensitivity function of contrast matching of sine-wave gratings. In these experiments, reduced sensitivity was noted for lower spatial frequencies but not for higher spatial frequencies. Our aim and method, however, were different: we adapted observers only to the spatial frequencies underlying a percept and observed the subsequent effects on perception; we did not adapt observers to the percept itself.

It is a challenge, however, to apply frequency-specific adaptation to figurative stimuli because such stimuli comprise many spatial frequencies at different amplitudes, orientations, and phases, most of which must be adapted to obtain the desired effect. The results of experiment 1 provide a complete description of the spatial frequencies that must be adapted to selectively affect the percept of the nuns or Voltaire (we describe the adaptation stimuli in section 3.1.2).

In the adaptation phase, one group of participants adapted to high-contrast dynamic noise created from the high spatial frequencies driving the perception of the nuns. In the other group, participants adapted to high-contrast dynamic noise created from the low spatial frequencies driving the perception of Voltaire. In each group, the coloured noise ensured an adaptation to the specific spatial frequencies underlying the percepts, but not to the percepts themselves. In a transfer phase, both participant groups were exposed to a low-contrast version of an ambiguous hybrid image composed of the information of both the nuns and Voltaire derived in experiment 1 (see figure 4). We expected the groups to experience orthogonal perceptions of this ambiguous image. Specifically, the group adapting to low spatial frequencies should perceive the nuns, and the group adapting to the high spatial frequencies should see Voltaire. These orthogonal perceptions subsequent to frequency channel adaptation would establish a possible link between the perceptual information isolated in experiment 1 and fundamental mechanisms of early vision.

3.1 *Method*

3.1.1 *Participants.* Participants were ten students from the University of Glasgow with normal, or corrected-to-normal, vision, who had not taken part in the previous experiment. Half of the participants were assigned to the Voltaire adaptation group and half to the nuns adaptation group.

3.1.2 *Stimuli.* The experiment ran on a Macintosh G4 with a program written with the Psychophysics Toolbox for Matlab (Brainard 1997; Pelli 1997) and the Pyramid Toolbox (Simoncelli 1997). For adaptation, we generated 200 white-noise fields, 256×256 pixels each (subtending 13.69 deg $\times$ 13.69 deg on the screen) that were filtered to contain the frequency-response profile (for the high-frequency noise, between 32 and 64 cycles per image; for the low-frequency noise, it was between 8 and 32 cycles per image) of each potent image derived from the previous experiment. Consequently, the noise patterns inherited their high contrast and random phases from the original white-noise stimuli—two important properties for frequency-specific adaptation.

The noise patterns were presented dynamically (ie in a movie) on a monitor with a screen refresh rate of 75 Hz on a black background. The presentation lasted for a total of 3 min 20 s. We constructed a post-adaptation hybrid image combining the low-contrast potent 'nuns' and the low-contrast potent 'Voltaire', a requirement of frequency-adaptation studies (see figure 5).

3.1.3 *Procedure.* Before experimentation, to ensure participants could perceive both interpretations of the Dali painting, they viewed the original grey-scale image from the first experiment. Participants were instructed to adapt to the moving noise patterns by moving their eyes over the pattern (Blakemore and Campbell 1969). A chin-rest maintained viewing distance at 40 cm. Immediately after adaptation, a low-contrast hybrid image comprising the potent 'nuns' and the potent 'Voltaire' was presented in the centre of the screen for 1 s. Participants were instructed to say aloud which of the two images they could perceive, the nuns or Voltaire.

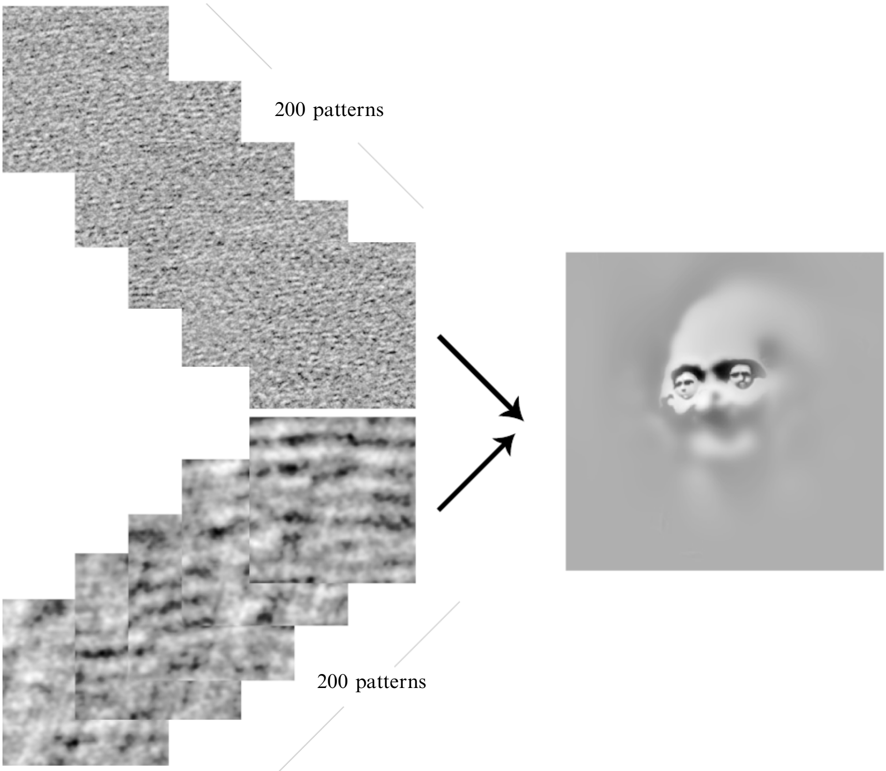

Figure 5. The stimuli and design of the adaptation experiment. High-spatial-frequency noise with the same amplitude as the nuns is depicted at the top and low-spatial-frequency noise with the same amplitude as Voltaire is shown at the bottom. Two hundred of these randomly generated noise patterns were repeated seventy-five times in a movie sequence. The phase was randomly disrupted. The rightmost picture is a hybrid comprising the potent nuns and Voltaire information (see figure 4) derived from the first experiment. The contrast of this image was reduced for post-adaptation.

3.2 *Results and discussion*

Of the ten participants, nine experienced an orthogonal perception, that is all of the subjects in the nuns-adaptation group who adapted to higher spatial frequencies perceived Voltaire on presentation of the hybrid, and, with the exclusion of one subject, those who adapted to low-frequency noise perceived the nuns. On applying a sign test the results proved to be significant ($p < 0.05$).

These results demonstrate a possible grounding in the mechanisms of early vision of the information underlying the selective perception of the nuns and Voltaire isolated in experiment 1 with 'bubbles'. Together, the results of experiments 1 and 2 have implications for the study of recognition and perception. An important issue pertains to the information content of a stimulus and its role in driving these processes. Here, we showed that the reversal between the two percepts in an ambiguous image could be understood in terms of the different scale information underlying these percepts. We believe that knowing which information people attend to, to resolve visual tasks, is an important but neglected component of perception. As we have demonstrated, the 'bubbles' technique can reveal what this information is.

4 Concluding remarks

In the first experiment, using an adapted version of the Gosselin and Schyns (2001) 'bubbles' technique to resolve a subjective task, we disambiguated the information for perception and suggested that the reversal between the two image percepts is the result of a switch in the spatial filters 'looking at' the image. In the second experiment, participants adapted to low-spatial-frequency or high-spatial-frequency noise to block the response of that frequency channel and henceforth, upon presentation of a hybrid image, experienced an orthogonal perception. Theoretically, we highlight the importance of understanding the information contained in a stimulus and how the nature of this information influences the perception process.

Acknowledgments. Thanks to the anonymous reviewers whose comments helped improve this manuscript. We would also like to thank Paul Sowden at the University of Surrey whose comments helped improve an earlier version of this manuscript. This work was supported by an ESRC grant (number R00223179) awarded to Philippe G Schyns.

References

Attneave F, 1971 "Multistability and perception" *Scientific American* **225**(6) 62 – 71

Blakemore C, Campbell F W, 1969 "On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images" *Journal of Physiology* **203** 237 – 260

Brainard D H, 1997 "The psychophysics toolbox" *Spatial Vision* **10** 433 – 436

Descharnes R, 1972 *The World of Salvador Dali* (London: MacMillan)

De Valois R L, De Valois K K, 1990 *Spatial Vision* (New York: Oxford University Press)

Gosselin F, Schyns P G, 2001 "Bubbles: A new technique to reveal the use of visual information in recognition tasks" *Vision Research* **41** 2261 – 2271

Hayes A, Ross J, 1995 "Lines of sight", in *The Artful Eye* Eds R Gregory, J Harris, P Heard, D Rose (Oxford: Oxford University Press) pp 339 – 352

Livingstone M, 2000 "Is it warm? Is it real? Or just low spatial frequency?" *Science* **290** 1299

Morrisson D, Schyns P G, 2001 "Usage of spatial scales for the categorization of faces, objects and scenes" *Psychological Bulletin and Review* **8** 454 – 469

Pantle A, Sekuler R, 1968 "Size-detecting mechanisms in human vision" *Science* **162** 1146 – 1148

Pelli D G, 1997 "The VideoToolbox software for visual psychophysics: Transforming numbers into movies" *Spatial Vision* **10** 437 – 442

Oliva A, Schyns P G, 1997 "Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli" *Cognitive Psychology* **34** 72 – 107

Schyns P G, 1998 "Diagnostic recognition: Task constraints, object information and their interactions" *Cognition* **67** 147 – 179

Schyns P G, Oliva A, 1999 "Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations" *Cognition* **69** 243 – 265

Simoncelli E P, 1997 "Image and multi-scale pyramid tools [Computer software]", available from the author: www.cns.nyu.edu/~eero/software.html

Webster M A, Miyahara E, 1997 "Contrast adaptation and the spatial structure of natural images" *Journal of the Optical Society of America A* **14** 2355 – 2366