Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 44 (2004) 471–477
Vision Research
www.elsevier.com/locate/visres

No troubles with bubbles: a reply to Murray and Gold

Frédéric Gosselin [a,*], Philippe G. Schyns [b]

[a] Département de Psychologie, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, Qué., Canada H3C 3J7
[b] Department of Psychology, University of Glasgow, 56 Hillhead Street, Glasgow G12 8QB, UK

Received 6 October 2003

Abstract

Murray and Gold discuss two "shortcomings" of the *Bubbles* method [*Vision Research* 41 (2001) 2261]. The first one is theoretical: *Bubbles* would not fully characterize the LAM (Linear Amplifier Model) observer, whereas reverse correlation would. The second "shortcoming" is practical: the apertures that partly reveal information in a typical *Bubbles* experiment would induce atypical strategies in human observers, whereas the additive Gaussian white noise used by Murray and Gold (and others) in conjunction with reverse correlation would not. Here, we show that these claims are unfounded.

1. Introduction

In the first part of their article, Murray and Gold formalize Gosselin and Schyns' (2002a) analysis of the relationship between the three main types of information involved in visual tasks, the $R \otimes A \approx P$ framework. We suggested that Potent information (P, rendered by *Bubbles*) is the result of an interaction (denoted by $\otimes$) of Represented information (R, measured by reverse correlation) with Available input information (A). Murray and Gold essentially confirmed that this analysis is valid in the context of the LAM observer and for a particular version of the *Bubbles* method.

We subscribe to the idea that without an analysis based on a precise observer model (such as LAM), it is difficult to know exactly what *Bubbles* and reverse correlation measure. Although we welcome Murray and Gold's formal effort, we do not believe that "Tout est pour le mieux dans le meilleur des mondes.", [1] as Voltaire's *Candide* famously pointed out. Therefore, before describing the LAM formalization and evaluating its real implications for *Bubbles*, we will first discuss its general limitations.

2. Beyond photometric space: general problems with the LAM observer

Experiments with reverse correlation and *Bubbles* estimate the visual information used in a task by testing how information from an image generation space modulates response. This space should be chosen with great care, because its structure will constrain the information-use estimates. Factors such as the nature of the task to be resolved and the nature of the stimuli should guide the choice. To appreciate the range and diversity of possible search spaces, consider the problem of estimating the features of face recognition when the stimuli are a few static 2D images. If the faces are normalized for the positions of their main features (e.g., eyes, nose, mouth, and silhouette) then the 2D photometric image space can be fruitfully searched, but the information estimates will be restricted in scope—essentially revealing attention to features in the 2D image. Suppose now that stimuli are 3D-rendered, laser-scanned faces. A linear parametric space might be used (e.g., morphing), with loss of linearity in the 2D image projections (e.g., Leopold, O'Toole, Vetter, & Blanz, 2001), but with greater scope for the information-use estimates—essentially transformations of the 3D shape of faces.

Image generation spaces constitute a pressing research issue because the structure of visual information will tend to change with tasks (e.g., identification, gender,

* Corresponding author.
E-mail addresses: frederic.gosselin@umontreal.ca (F. Gosselin), philippe@psy.gla.ac.uk (P.G. Schyns).

[1] "All is for the better in the best of worlds."

0042-6989/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2003.10.007

expression, age, and so forth) and object classes (e.g., static 2D pictures, static 3D-rendered objects, dynamic objects and so forth, see Schyns, 1998). In our view (and that of others, e.g., Mangini & Biederman, in press; Ollman & Kersten, in press; Sadr & Sinha, 2003), recognition researchers should not restrict the scope of their research to the linear photometric space of the LAM observer, for a number of straightforward reasons discussed below.

Reverse correlation of the observer's responses onto Gaussian white noise of the type advocated by Murray and Gold does *not* provide a truthful representation of the information that the observer could use in the task. Instead, this version of reverse correlation projects the used information onto a linear space, as a weighted sum of Gaussian white noise pixel intensities. Such linear estimates have paid off particularly well in neurophysiological and psychophysical studies of lower level vision, where intensity-based receptive fields are responsive to noise induced modulations of input contrast (Ringach & Shapley, in press; for a review). The correlation between added input noise and response magnitude guarantees a good linear regression.

The success of models of low-level vision does not necessarily entail success of the same models at higher levels of vision. The neural systems of face and object categorization do not respond with the same magnitude to simple modulations of contrast like low-level ones do (Avidan, Harel, Ben-Bashat, Zohary, & Malach, 2001), but rather they respond to scale, higher level features and shape information (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Haxby, Hoffman, & Gobbini, 2000; Kanwisher, McDermott, & Chun, 1997; Op de Beeck, Wagemans, & Vogels, 2001; Sigala & Logothetis, 2002).

In sum, LAM is not sufficiently general in scope to impose any prescriptive standards on the conduct of research in visual categorization (see Gosselin & Schyns, in press). Thus, Murray and Gold's critique of *Bubbles* should be taken with a grain of salt. Not only is it limited because it assumes that humans are LAM observers—which they are not, but also because it assumes a particular version of *Bubbles*. This last point is the object of the next section.

3. Within photometric space: limitations in scope of the Murray and Gold analysis

Bubbles is an information sampling technique in which a "bubble" carves out an information sample from the generation space. With the exception of Schyns and Gosselin (2002), we have so far restricted the application of *Bubbles* to photometric spaces, not all of which have a 2D structure. [2] However, *Bubbles* is a generic method that can be applied to more complex parametric spaces than the photometric space (see Gosselin & Schyns (in press) for discussion), although the same also applies to reverse correlation (see Ollman & Kersten, in press). Murray and Gold's analysis applies only to the low-passed white noise version of *Bubbles* in photometric space. We will now show that this analysis with limited scope does not demonstrate the superiority of reverse correlation over *Bubbles* for the practice of everyday research.

4. Still RAPing

The key formal development that Murray and Gold provide is reproduced below (their Eq. (5)):

$$B \approx b * b * (T \circ (I_X - I_Y)).$$

For a LAM observer, T (or R, in our *RAP* framework) is an internal template, $(I_X - I_Y)$ (or A) is the ideal template, B (or P) the *Bubbles* information estimate, b the Gaussian bubble used to low-pass the white noise window, $\circ$ a pointwise product, and $*$ a convolution. This formulation implements $R \otimes A \approx P$ in the context of a LAM observer: (1) the fuzzy $\otimes$ operator becomes $\circ$, and (2) $(b * b)$, is introduced. The latter is a "double-blurring" term that quantifies the limit of the spatial resolution that can be achieved with a particular Gaussian b, given σ. It is easy to show that $b(\sigma = s) * b(\sigma = s)$ reduces to $b'(\sigma = \sqrt{2}s)$, a larger single Gaussian bubble.

In *Bubbles*, σ is the only parameter that must be adjusted to sample information in a 2D image space. We have already shown that standard deviations of different sizes can be simultaneously used to search the image (e.g. Gosselin & Schyns, 2001a, 2001b; Schyns et al., 2002). The choice of an appropriate σ is subject to a number of parameters (not all independent) ranging from the expected scale of visual information, the required smoothness of the solution, the number of parameters to estimate, and the required rate of convergence. *Bubbles* solutions can range from coarse (i.e., with a large Gaussian bubble, few parameters to estimate, and typically fast convergence) to fine (i.e. with many parameters to estimate and typically slow convergence). At the limit, sigma becomes a dot in discrete space, and $b * b$ vanishes from Murray and Gold's equation, which then reduces to $R \circ A \approx P$, and the

[2] A list of generation spaces explored with *Bubbles*: the standard 2D image plane (Gibson, Gosselin, Wasserman, & Schyns, submitted; Gosselin & Schyns, 2001a; Schyns, Jentzsch, Johnson, Schweinberger, & Gosselin, 2003; O'Donnell, Schyns, & Gosselin, 2002), spatial frequency (Schyns & Gosselin, 2002), spatial frequency × 2D image (Bonnar, Gosselin, & Schyns, 2002; Gosselin & Schyns, 2001a; Schyns, Bonnar, & Gosselin, 2002), and 2D image plane × time (Vinette, Gosselin, & Schyns, in press).

spatial resolution of *Bubbles* is equivalent to that which is typical of white noise reverse correlation.

5. No troubles in theory: LAMe attack

The Murray and Gold formalization led them to conclude: "[...] if the LAM is a valid model of the system under study, then [...] Eq. (5) shows that from a classification image [that is R] we can easily determine the result of any bubbles experiment. [...] Thus in cases where the LAM is correct, the bubbles method is *superfluous* [...]" (p. 9, italic added). Later on (on p. 19), they modulate this statement with: "[...] if one is interested only in what stimulus locations help an observer give a correct response [that is P], then the bubbles method is perfectly adequate [...]".

We have shown that Murray and Gold's equation (5) reduces to $R \circ A \approx P$ when σ is so small that the Gaussian bubble reduces to a dot. Whenever $R \circ A \approx P$ and $R \approx P/A$ (a pointwise division) are defined, a complete *RAP* characterization of the observer can be obtained either from A and R (estimated from reverse correlation) or from A and P (estimated from *Bubbles*). The theoretical superiority of reverse correlation over *Bubbles* rests solely on the domain of definition: $R \circ A$ is defined over all real numbers, whereas P/A is undefined whenever $A(x,y) = 0$, for any x and y. To make it absolutely clear, reverse correlation is superior to *Bubbles* whenever there is no information available at a given image location to resolve a task—whenever $A(x,y) = 0$.

In our own research, we used "superstitious" to refer to situations in which $A(x,y) = 0$ and $R(x,y) \neq 0$. For example, we induced three observers to "superstitiously" see an 'S' letter ($R(x,y) =$ 'S') in pure bit noise ($A(x,y) = 0$) by artificially restricting (via instructions) the number of candidate representations to be matched against the input noise. We then applied reverse correlation to depict the observer's share, i.e. $R(x,y)$. We are thus well aware that reverse correlation can—and that *Bubbles* cannot—be applied in such "superstitious" situations. We made this point explicit a number of times (Gosselin, Bacon, & Mamassian, submitted; Gosselin & Schyns, 2002a, 2002b; Gosselin & Schyns, 2003a, 2003b).

Unless artificially created, [3] however, $A(x,y) = 0$ is the exception. In most of the work by Murray and Gold and colleagues, two-image experiments are designed. To derive A in such conditions, we computed the differences between all possible pairs of the 32 face images (grayscale 8-bits) used in Gosselin and Schyns (2001a, Experiment 1). $A(x,y) = 0$ occurs with a probability of 0.022, always within the forehead region. [4] In most recognition experiments (and real life), observers compare pairs of categories of images, not pairs of individual images. In LAM, the information available between pairs of categories is the sum of the images in one category minus the sum of the images in the other category. To set the stage for the experiment to come, columns "A" in Figs. 1 and 2, respectively, present the information available in the GENDER (male face $1 + \cdots +$ male face 10−female face $1 - \cdots -$ female face 10) and expressive or not (EXNEX) categorizations (smiling face $1 + \cdots +$ smiling face 10−neutral face $1 - \cdots -$ neutral face 10) computed from the 20 8-bit grayscale images of Schyns et al. (2002). The likelihood of $A(x,y) = 0$ is .00034 in GENDER and .000057 in EXNEX. As human observers tend to maximize within-category similarity and between-category dissimilarity (Gosselin & Schyns, 2001b; Rosch, 1978), the closer experimental stimuli get to real-world conditions of stimulation, the smaller these probabilities will be.

If $A(x,y) = 0$, and an estimate of $R(x,y)$ is required from $P(x,y)/A(x,y)$, then 0 can be assigned to $R(x,y)$. If there is no information available for this pixel, it is reasonable to assume that the observer's representation will not comprise information for this pixel either.

To summarize, the fact that $P(x,y)/A(x,y)$ is not defined when $A(x,y) = 0$ gives little support to the claim that the *Bubbles* method is "superfluous" (Murray and Gold, p. 9, 19). We have argued that $A(x,y) = 0$ is an outlier event in the real-world, that it only occurs in artificial laboratory situations. For most practical purposes, then *Bubbles* and reverse correlation are complementary techniques (see Gosselin & Schyns, 2002a).

6. No troubles in practice: a fair comparison of bubbles and reverse correlation

To compare the outcomes of reverse correlation and *Bubbles*, we applied reverse correlation in two categorization tasks (GENDER, and expressive or not, EXNEX) in experimental conditions identical to the *Bubbles* experiment of Schyns et al. (2002). Fifteen University of Glasgow paid observers were assigned to each categorization task (a total of 30 observers). A trial consisted in the presentation of a randomly chosen stimulus to which Gaussian white noise was added. The

[3] Murray and Gold (p. 17) "[...] chose the fat–thin task [...] in order to show how the bubbles method would affect strategies in a task that had actually been discussed in the literature, rather than a task that was designed to maximize the disruptive effect of showing only small fragments of a signal." Unfortunately, in this fat–thin task, A has many more zero information pixels than non-zero ones—non-zero information pixels of A are concentrated in two sectors of 3.5° at the palates of each "pacman". As this is a most unfortunate choice of experiment to compare *Bubbles* with reverse correlation we will not discuss Murray and Gold's experiment any further.

[4] We did the same for the 10 face images (grayscale 8-bits) of Gold, Bennett, and Sekuler (1999a, 1999b). $A(x,y) = 0$ never occurs.

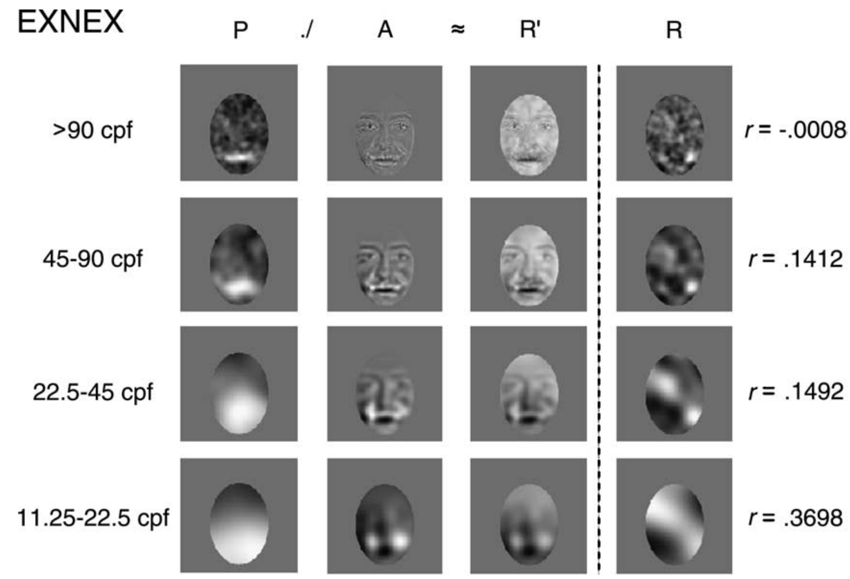

Fig. 1. This figure depicts the outcomes of the EXNEX task in the $R \otimes A \approx P$ framework (Gosselin & Schyns, 2002a). P and R, respectively, depicts the *Bubbles* and the reverse correlation estimates of visual information at different scales (R is "double-blurred"). A depicts the linear information available to resolve the EXNEX categorization task. Lastly, the figure depicts R' that is, the predicted R (R' is the best linear fit of P/A to R). The Pearson correlations give the similarity between R and R'.

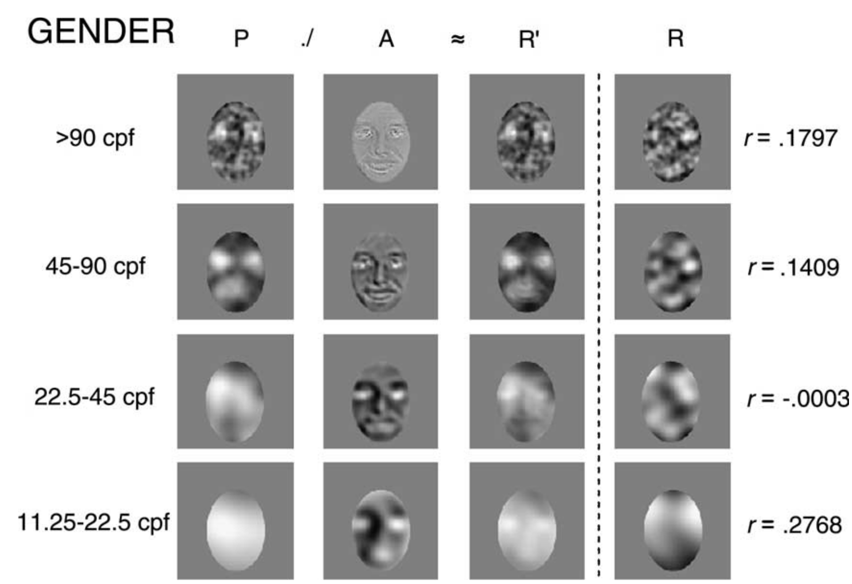

Fig. 2. This figure depicts the outcomes of the GENDER task in the $R \otimes A \approx P$ framework (Gosselin & Schyns, 2002a). P and R, respectively, depicts the *Bubbles* and the reverse correlation estimates of visual information at different scales (R is "double-blurred"). A depicts the linear information available to resolve the GENDER categorization task. Lastly, the figure depicts R' that is, the predicted R (R' is the best linear fit of P/A to R). The Pearson correlations give the similarity between R and R'.

sigma of the noise distribution was adjusted to maintain observers' performance at 75% correct. The stimuli were presented on a calibrated high-resolution Sony monitor, with a refresh rate of 85 Hz. The *Bubbles* version of the experiment differed only in how the facial information was sampled. For the reverse correlation experiment, results were derived from the pooled data of 15 observers each resolving 1000 trials, separately for the GENDER and EXNEX conditions. For the *Bubbles* experiment, we reanalyzed the data of Schyns et al. (2002), pooling this time 1000 trials per observer, separately for GENDER and EXNEX (the analyses reported in Schyns et al., 2002 only concerned the last 500 trials of each observer).

The raw *Bubbles* estimates, at different scales, in the GENDER and EXNEX tasks are shown in columns

"*P*" of Figs. 1 and 2 (see Schyns et al. (2002), for details). We computed the raw classification images by linearly combining the Gaussian noise fields according to the categorization response they elicited (i.e. hits + false alarms−misses−correct rejections).

To compare the reverse correlation solution to the *Bubbles* solution we convolved the former with Gaussians of standard deviations 8.48, 16.97, 33.94 and 67.88 pixels, from fine to coarse (this takes into account the "double-blurring" discussed by Murray and Gold). The outcome is shown in columns "*R*" of Figs. 1 and 2. To facilitate comparisons between *Bubbles* and reverse correlation, we computed $P/A = R'$ (see columns "R'" in Figs. 1 and 2). R and R' are directly comparable. At most scales, R and R' are positively correlated. While the *Bubbles* outcome (and this also applies to R') remains consistent at coarser scales, the corresponding reverse correlation solutions become noisier and more difficult to interpret.

In sum, it does not appear that "...the *Bubbles* method drastically changes human observers' strategies," (Murray and Gold, p. 4) at least compared with Gaussian noise reverse correlation. Furthermore, there is no evidence that the value of the information extracted from the *Bubbles* experiment is reduced compared to that extracted from the reverse correlation experiment.

7. No troubles whatsoever in practice

Even though we did not find empirical evidence for this, we are still left with Murray and Gold's firm belief that the type of windowing used in *Bubbles* "drastically" disrupts the observers' strategy, whereas additive Gaussian white noise does not (Murray and Gold, Abstract, p. 4, 12, 17, 19, and 20). [5] They develop three arguments to support their belief. None of them survives closer scrutiny.

Murray and Gold wrote (p. 16):

> First, it is intuitively clear why windowing stimuli through bubbles might change observers' strategies: when only small parts of a stimulus are shown on any given trial, observers may be forced to use stimulus features that they would not use if the whole stimulus was presented.

To the extent that added Gaussian white noise will mask certain diagnostic regions more than others, observers that would have used the former will have to resort to using the latter. The probability that a diagnostic region will be entirely destroyed by additive noise might appear small in comparison to the probability that it is not revealed through Gaussian bubbles. However, this will happen only if the sigma of the Gaussian bubbles is too large for the task at hand (see discussion in Section 4). This does not demonstrate the superiority of additive noise over windowing.

> [Second,] a great deal of psychophysical and physiological evidence shows that even under noiseless viewing conditions, observers' performance in threshold tasks is limited by internal noise, so by adding external noise we are probably not presenting observers with a task that is qualitatively different from a noiseless threshold task [...]. (Murray and Gold, p. 16–17)

This could be the case, but the critical part of this argument (i.e., that observers are not performing a task qualitatively different from an everyday visual task) applies to windowing as well. Are not the eyes a window on the world? And are not objects almost never seen in their entirety? As Murray, Sekuler, and Bennet (2001) put it: "One of the challenges to object recognition is the fact that sensory information reaching the eyes is often incomplete: Objects occlude parts of neighboring objects and parts of themselves. Even though we constantly perceive partly occluded objects, we rarely notice that the visual information we receive is incomplete." (p. 1) Thus, windowing is certainly a type of noise we are accustomed to dealing with in the real-world.

> Third, and most convincingly, observers' contrast energy thresholds have been found to be an approximately linear function of external noise power in practically every task in which this relationship has been tested, including discrimination of fat vs. thin Kanizsa squares, and this is strong evidence that observers use the same strategy at all levels of external noise, from negligible levels to high levels of noise. (Murray and Gold, p. 17)

An analoguous approximate linear relationship was found between contrast energy threshold and the area revealed by Gaussian bubbles. Two experienced psychophysical observers from the Université de Montréal with normal, or corrected to normal vision, resolved either the GENDER (EM), or the EXNEX (IF) on the face set employed in the experiment described in Section 6 (see also Schyns et al. (2002)). Observers had previously participated in several other *Bubbles* experiments with the same face set and with the same tasks; their

[5] As demonstrated by Murray and Gold, a LAM observer uses the same strategy when presented with bubbled stimuli than when presented with stimuli plus Gaussian white noise. Therefore any difference between the strategies induced by windowing with Gaussian bubbles and additive noise on human observers refutes LAM as an adequate model. The success of Murray and Gold's practical argument entails the demise of their theoretical argument, and vice-versa.

performance had thus stabilized at the time of this investigation. The experiment ran on a Macintosh G4 computer using a program written with the Psychophysics Toolbox for Matlab (Brainard, 1997; Pelli, 1997). The stimuli were presented on a calibrated high-resolution Sony monitor, with a refresh rate of 85 Hz. Contrast energy thresholds for a 75% correct response rate were measured at four levels of windowing (400, 500, 600, and 700 bubbles with a standard deviation equal to 3 pixels, or 0.067° of visual angle, for stimuli spanning 256×256 pixels, or $5.72 \times 5.72°$ of visual angle) using the method of constant stimuli (we employed five levels of contrast energy for each level of windowing – a total of $96 * 5$ levels of contrast energy $* 4$ levels of windowing $= 1920$ trials per observer; R^2 for the best-fitted cumulative Gaussians ranged from 0.692 to 0.947, with an average of 0.803). We found approximate linear relationships between contrast energy thresholds and levels of windowing (IF: $R^2 = 0.947$; EM: $R^2 = 0.738$).

In sum, the Murray and Gold claim that use of *Bubbles* "drastically" modifies observers' strategies and so reduces the value of the technique compared with additive Gaussian white noise reverse correlation is unfounded. Our face recognition comparison did not lend support to this claim. In addition, *none* of the arguments put forth by Murray and Gold claiming that Gaussian white noise is preferable to Gaussian bubbles stands up to closer scrutiny.

8. Conclusions

Murray and Gold claimed that there are some "shortcomings" with *Bubbles*. We have addressed their criticisms and shown: (1) that their formal analysis is restricted in scope; (2) that the argument that *Bubbles* would not fully characterize the LAM observer is inconsequential; (3) that in a fair comparison, *Bubbles* and reverse correlation reveal a similar use of information in human observers; and (4) that none of the arguments put forth for preferring additive Gaussian white noise over windowing with Gaussian bubbles survives closer scrutiny. Thus, there are no troubles with *Bubbles*.

Acknowledgements

This research was supported by a NSERC (R0010085) and a NATEQ (R0010287) grant awarded to Frédéric Gosselin; and by an ESRC grant (R000237901) awarded to Philippe Schyns. We thank Marie Smith for running the first experiment, and Éric McCabe and Isabelle Fortin for their help with the second experiment. Thanks to Marie Smith and Asifa Majid for their comments on the article.

References

Avidan, G., Harel, M., Ben-Bashat, D., Zohary, E., & Malach, R. (2001). Contrast sensitivity in human visual areas and its relationship to object recognition. *Journal of Neurophysiology, 87*, 3112–3116.

Bonnar, L., Gosselin, F., & Schyns, P. G. (2002). Understanding dali's slave market with the disappearing bust of voltaire: A case study in the scale information driving perception. *Perception, 31*, 683–691.

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

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience, 2*, 568–573.

Gibson, B. M., Gosselin, F., Wasserman, E. A., & Schyns, P. G. (submitted). Pigeons and people categorize human faces using similar features.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999a). Signal but not noise changes with perceptual learning. *Nature, 402*, 176–178.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999b). Identification of band-pass filtered faces and letters by human and ideal observers. *Vision Research, 39*(21), 3537–3560.

Gosselin, F., Bacon, B. A., & Mamassian, P. (submitted). Internal surface representations approximated by multiple regression. *Vision Research*.

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

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

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

Gosselin, F., & Schyns, P. G. (2002b). You are about to see picturial representations! [Commentary on Z. Pylyshyn's target article]. *Behavioral and Brain Research, 25*, 191–192.

Gosselin, F., & Schyns, P. G. (2003a). A user's guide to bubbles. In L. Gershkoff-Stowe & D. Rakison (Eds.), *Building categories in developmental time.* Hilldsale, NJ: Erlbaum (in press).

Gosselin, F., & Schyns, P. G. (2003b). Superstitious perceptions reveal properties of memory representations. *Psychological Science, 14*, 505–509.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*, 223–233.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate visual cortex specialized for face perception. *Journal of Neuroscience, 17*, 4302–4311.

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level after effects. *Nature Neuroscience, 4*, 89–94.

Mangini, M., & Biederman, I. (in press). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science*.

Murray, R. F., Sekuler, A. B., & Bennet, P. J. (2001). Time course of amodal completion revealed by a shape discrimination task. *Psychonomic Bulletin and Review, 8*, 713–720.

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

Ollman, C. A., & Kersten, D. (in press). Classification objects, ideal observers and generative models. *Cognitive Science*.

Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parametrized shapes. *Nature Neuroscience, 4*, 1244–1252.

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

Ringach, D., & Shapley, R. (in press). Reverse correlation in neurophysiology. *Cognitive Science*.

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

Sadr, J., & Sinha, P. (2003). Object recognition and random structure evolution. *Cognitive Science*.

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

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

Schyns, P. G., & Gosselin, F. (2002). Use of frequency information at the basic and the subordinate level in animal categorization. *Vision Sciences*.

Schyns, P. G., Jentzsch, I., Johnson, M., Schweinberger, S. R., & Gosselin, F. (2003). A principled method for determining the functionality of brain responses. *NeuroReport*.

Sigala, N., & Logothetis, N. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature, 415*, 318–320.

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

Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Vision Research 44 (2004) 479–482
Vision Research
www.elsevier.com/locate/visres

Reply to Gosselin and Schyns

Richard F. Murray [a,*], Jason M. Gold [b]

[a] Center for Perceptual Systems, University of Texas at Austin, 1 University Station A8000, Austin, TX 78712-0187, Austin, USA
[b] Department of Psychology, Indiana University, 1101 East 10th Street, Bloomington, IN 47405, USA

Received 29 September 2003

Abstract

We discuss Gosselin and Schyns' (2003) reply to our criticisms and constructive suggestions concerning the bubbles method Murray and Gold (2003). We find that their reply does not mollify our concerns, and we still believe that reverse correlation will generally be preferable to the bubbles method until further developments (a) demonstrate more clearly what the bubbles method actually measures and (b) introduce a type of windowing noise that is less likely to disrupt observers' strategies.

1. Our LAM analysis of bubbles is too limited to be of any value

Gosselin and Schyns' reply does not alleviate our concerns about the bubbles method in its present form. Here we discuss their most important claims, and explain why we do not think that they adequately address our criticisms and constructive suggestions.

Gosselin and Schyns argue that "the LAM is not sufficiently general in scope to impose any prescriptive standards on the conduct of research in visual categorization", and hence that our LAM-based analysis is not useful. Certainly the LAM is an incomplete model of human performance, but this sweeping judgement is far too dismissive. As we said in our article, the LAM is a useful first-order approximation that captures many aspects of human performance, and serves as a starting point for more complex models. Furthermore, many nonlinear models are locally linear, which means that a linear analysis is often adequate in psychophysical tasks where the stimuli cover only a narrow range (Ahumada, 1987). Even in high-level tasks like face and letter identification, an analysis of observers' performance in terms of templates and internal noise can lead to robust and surprising results (e.g., Gold, Bennett, & Sekuler, 1999; Solomon & Pelli, 1994; Tjan, Braje, Legge, & Kersten, 1995).

Furthermore, the LAM and our LAM-based analysis are not as restrictive as Gosselin and Schyns imply. They claim that our analysis can only be applied to tasks in which the bubbles window small spatial regions, rather than windowing regions in some more abstract representation of the stimulus, such as a scale space. This is simply not true. We described the bubbles method in terms of spatial bubbles, because this is the approach that Gosselin and Schyns have used in almost all their work, but our analysis used a very general and abstract description of the bubbles method. Our analysis can be used whenever the observer's decision variable is regarded as a linear, Gaussian-noise contaminated function of the stimulus in *some* representation, however abstract. (In particular, the representation could be related to the photometric representation by the arbitrarily complex morphing operations mentioned by Gosselin and Schyns.)

In any case, our LAM-based analysis is the *only* rigorous analysis of the bubbles method to date. Gosselin and Schyns criticize our approach as being too simplistic, but as we made clear, we believe that it is a limited but useful first step in putting their method on a rigorous footing. For instance, we showed that their RAP law is valid for linear observers, and this led us to note that in general, for nonlinear observers, it is *not* valid—surely a useful contribution, as they themselves often make use of their RAP law, but have never discussed either its justification or its domain of validity.

* Corresponding author. Address: Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, PA 19104-6196, USA. Tel.: +1-215-898-7300; fax: +1-215-746-6848.
E-mail address: murray@psych.upenn.edu (R.F. Murray).

0042-6989/$ - see front matter
doi:10.1016/j.visres.2003.10.008

Furthermore, we did suggest that for anyone interested in using the bubbles method, it is important to extend our LAM analysis to more complex models. Thus our credo is not the younger, naïve Candide's "*tout est pour le mieux*", but his later conclusion after long experience that "*il faut cultiver notre jardin*" [1]—in this case, a garden of bubbles.

2. A bubbles image completely reveals an observer's template

Gosselin and Schyns accept our proof that for a LAM observer, the expected value of a bubbles image is given by our equation (5):

$$E[B] = u + v \cdot b * b * (T \circ (I_X - I_Y))$$

They argue, nevertheless, that a bubbles image completely recovers a LAM observer's template in all interesting cases. They suggest that one can recover the template by using a single-pixel bubble, so that the double-convolution disappears, and then by dividing the bubbles image pointwise by the ideal template, $I_X - I_Y$, leaving a term proportional to the template T except at points where $I_X - I_Y = 0$, at which locations the division is undefined.

We have two initial objections that are important to note, but that can be met. First, the term u must first be subtracted from the bubbles image for this scheme to work. It can be calculated easily; see our Appendix (Murray & Gold, 2003). Second, no bubbles experiment has ever been carried out with single-pixel bubbles, and several of the reported experiments actually rely on using several sizes of bubbles, so it is not clear how usable this single-pixel scheme is in bubbles experiments as they are actually practised. However, even if a multi-pixel bubble is used, the effects of the double-convolution can be undone by deconvolution. The resulting signal-to-noise ratio will be low at high spatial frequencies, but nonetheless, with some effort this approach might be made to work.

Our main objection is that, contrary to Gosselin and Schyns' claims, even a cursory review of the reverse correlation literature shows that observers' responses are *often* influenced by stimulus locations where the ideal template is zero. Observers use irrelevant landmarks in vernier alignment tasks (Ahumada, 1996), illusory and occluded contours in shape discrimination tasks (Gold, Murray, Bennett, & Sekuler, 2000), irrelevant, uncued locations in attention tasks (Shimozaki, Eckstein, & Abbey, 2002), and uninformative pre- and post-stimulus intervals in detection tasks (Neri & Heeger, 2002). Furthermore, observers invariably use uninformative stimulus regions surrounding informative regions, probably because of spatial uncertainty. It is impossible to recover any of these regions of observers' templates with Gosselin and Schyns' scheme. Consequently, as we claimed, a bubbles image does not completely determine a LAM observer's template.

Moreover, a moment's reflection shows that Gosselin and Schyns' scheme for recovering a LAM observer's template from a bubbles image is ill-conditioned not only at stimulus locations where the ideal template is zero, but also where it is *near* zero. In a bubbles image, these two types of locations will generally have almost indistinguishably close values, because neither of them greatly help the observer to give a correct response. Dividing the bubbles image by near-zero locations in the ideal template will magnify the inevitable statistical noise enormously, and the resulting estimate of the template will be practically useless.

Thus, contrary to Gosselin and Schyn's claim, reverse correlation does recover much more information about LAM observers than the bubbles method does.

3. The bubbles method does not change observers' strategies

Gosselin and Schyns carry out a face identification experiment to compare the strategies that observers use in bubbles and reverse correlation experiments, and they conclude that the strategies are roughly similar. Their interpretation of this experiment is fatally flawed in two ways. First, Gosselin and Schyns assume that a LAM observer in their categorization task can have only a single template. In a categorization task with many possible stimuli and just two responses, a LAM observer is normally assumed to have a stored template for each possible stimulus (e.g., Peterson, Birdsall, & Fox, 1954; Tjan et al., 1995). In this case, the calculation of the classification image is difficult, and must be done separately for each possible stimulus-response pair (Watson, 1998). Thus Gosselin and Schyns miscalculate the classification image: one cannot simply sum the noise images within each response category, and take the difference of these sums, as they do.

Second, and more crucially, Gosselin and Schyns make a basic logical error. We claim that, in many tasks, the bubbles method will change observers' strategies, and we have shown that this is demonstrably true in at least one task (the fat–thin task). Thus we have shown that, in general, one must be concerned that the bubbles method may change observers' strategies. Gosselin and Schyns show that in one task the bubbles method (perhaps) does not change observers' strategies, but from this they cannot conclude that in general, it does not change observers' strategies. In fact, our single counterexample (the fat–thin task) shows that it sometimes *does*. Unfortunately, they do not discuss our experiment,

[1] We must cultivate our garden.

so we do not know why they think it is inconclusive. (They do not discuss our experiment with respect to the question of whether a bubbles image completely recovers an observer's template because they believe that there are too many zero-valued pixels in the ideal template for the fat–thin task. However, this objection has nothing to do with the question of whether the bubbles method changes observers' strategies.)

Finally, Gosselin and Schyns discuss the three theoretical reasons that we gave to support our claim that the bubbles method is more likely to change observers' strategies than reverse correlation.

(a) We argued that the obliteration of large, randomly chosen parts of the stimulus is more likely to make observers change their strategies from trial to trial, than is adding Gaussian white noise. They argue that bubbles are more disruptive than Gaussian noise only if the bubbles are too large. Our response is that (i) bubbles experiments to date have used a small number ($\sim$25) of large bubbles, so our objection was more than theoretical, and (ii) the idea of using many tiny bubbles is very similar to our suggestion that one should use multiplicative Gaussian white noise, rather than randomly placed bubbles, and we agree that this would be a useful modification to the bubbles method.

(b) We noted that many psychophysical and physiological experiments have shown that observers must contend with internal Gaussian noise, even when there is no external noise, and we suggested that moderate amounts of external noise are therefore unlikely to drastically change observers' strategies. They point out that it is also the case the parts of objects are often occluded, as in a bubbles experiment. However, occlusion seems usually to occur in the form of large, contiguous segments of objects being occluded by other objects. Do we normally identify faces through 25 small randomly placed holes in occluding surfaces? To us, the analogy seems more than a little strained.

(c) We noted that noise masking functions are typically linear, indicating that observers' sampling (i.e., template) efficiency is not drastically altered by adding external white noise. In reply, Gosselin and Schyns report a new face identification experiment showing that the threshold contrast energy of a (complete, pre-windowed) stimulus declines linearly as a function of the number of bubbles. Our response is that, first, it would be helpful to see a careful explanation of *why* this implies that observers' strategies are constant as a function of the number of bubbles. Essentially, this result shows that the on-screen contrast energy is constant as a function of the number of bubbles, and it is certainly plausible that this is the signature of a constant strategy. However, to take just one possible problem, it is not clear how an observer's uncertainty concerning the number and position of bubbles will complicate this picture. We cited the linearity of noise masking functions in our argument, because it can be shown to correspond to constant sampling efficiency as a function of external noise power (Burgess, Wagner, Jennings, & Barlow, 1981). No such results have ever been derived to aid in the interpretation of Gosselin and Schyns' experiment. Their interpretation is plausible, and we are willing to believe it, but it must be shown to be correct.

Our second objection is that in all bubbles experiments to date, Gosselin and Schyns have used around 25 medium-sized bubbles, whereas in this experiment they use 400–700 very small bubbles. In our view, the reason why the bubbles method changes observers' strategies is that when small, randomly chosen parts of the stimulus are shown from trial to trial, the observer will use whichever part is available on any given trial. When the stimulus is shown through a very large number of very small bubbles, the situation changes entirely: on any given trial, one or more bubbles are very likely to fall in any reasonably large stimulus region, and there will be less incentive for the observer to change his strategy from trial to trial. In fact, in the theoretical limit of an extremely large number of extremely small bubbles (e.g., the size of a monitor phosphor molecule), changing the number of bubbles is tantamount to simply changing the stimulus contrast. Thus the validation experiment was carried out under very different conditions than all bubbles experiments to date, and we think that it gives little support to the bubbles method as it has actually been used. (Again, the notion of using many tiny bubbles is very similar to our suggestion of using multiplicative white Gaussian noise, and we do agree that with this new modification, the bubbles method is less likely to disrupt observers' strategies.)

Finally, and most crucially, Gosselin and Schyns make the same fatal logical error as before: we have shown that in at least one task (the fat–thin task) the bubbles method does disrupt observers' strategies, and even if Gosselin and Schyns' interpretation of their face identification experiment is correct, citing one task where the bubbles method does not disrupt observers' strategies does little to refute the conclusion that, in general, one must be concerned that using the bubbles method in studying a novel task will disrupt observers' strategies. Again, they do not discuss our experiment, so we do not know why they think it is inconclusive.

4. Conclusion

We still believe that the bubbles method could be a useful addition to the current library of system identification methods. However, the problems with the method in its present form are serious, and they cannot be argued away. As we outlined in our article, we believe that the correct approach is to develop the method further, by analyzing it rigorously in the context of more sophisticated models of vision, and by experimenting with different forms of visual noise to find a way of minimizing the tendency of the bubbles method to disrupt observers' strategies.

References

Ahumada, A. J. (1996). Perceptual classification images from vernier acuity masked by noise. *Perception, 25*, 18.

Ahumada, A. J. (1987). Putting the visual system noise back in the picture. *Journal of the Optical Society of America A, 4*(12), 2372–2378.

Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science, 214*(4516), 93–94.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature, 402*(6758), 176–178.

Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioral receptive fields for visually completed contours. *Current Biology, 10*(11), 663–666.

Gosselin, F., & Schyns, P. G. (2003). No troubles with bubbles. *Vision Research, 44*.

Murray, R. F., & Gold, J. M. (2003). Troubles with bubbles. *Vision Research, 44*.

Neri, P., & Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature Neuroscience, 5*(8), 812–816.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory, 4*, 171–212.

Shimozaki, S., Eckstein, M. P. & Abbey, C. (2002). Uncertain humans in a structurally certain world: attentional leaking with 100% valid postcues as seen by classification images. Fall Vision Meeting, San Francisco, CA, October 24–27, 2002.

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature, 369*(6479), 395–397.

Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research, 35*(21), 3053–3069.

Watson, A. B. (1998). Multi-category classification: template models and classification images [ARVO abstract #4197]. *Investigative Ophthalmology and Visual Science, 39*(4), S912.