$R \quad A \approx P$

Frédéric Gosselin and Philippe G. Schyns

University of Glasgow

Submitted for publication

Please send all correspondence to:

Frédéric Gosselin or Philippe G. Schyns
Dept of Psychology
University of Glasgow
56, Hillhead Street
Glasgow G12 8QB
UK
{gosselif, philippe}@psy.gla.ac.uk
phone: 44 141 330 49 37

ABSTRACT

Cognitive Science is about several disciplines communicating their different perspectives on the mind, the common object of study. However, as the disciplines themselves often illustrate, domain-specific concepts and techniques can prevent, rather than foster, a communication of viewpoints. In this article, we develop $R \quad A \approx P$ a new framework suggesting that the interaction between Represented R information and Available A information determine the Potent information P. We argue and illustrate that this framework is useful to establish a common language to articulate issues common to low-, mid-, and high-level vision. More importantly, we present new techniques (*reverse correlation* and *Bubbles*) with which to visualize the so far elusive constructs of representation and potent information.

In Jaina metaphysics, interpreting experience from a single point of view is making an error comparable to that of the six blind men touching an elephant (e.g., Backstein, 1992). A first blind man runs straight into an elephant's broad smooth side and concludes that it is a wall; a second blind man pokes the animal's trunk and thinks that it is a snake; a third one walks into the elephant's tusk to conclude that it is a spear; a fourth blind touches one leg and infers that it is a mighty column; a fifth blind man holds the elephant's tail and believes that it is a frayed bit of rope; and the sixth blind man grasps one of the flapping ears and believes that it is a fan. Jaina doctrine recommends that we apprehend reality from a multiplicity of different points of view.

Cognitive Science is Jainian in essence: several sub-disciplines apply their own tinted lenses to the observation of the mind, the common object of investigation. The confrontation of several viewpoints should in principle minimize the distorting influence of any one of them, revealing the workings of the mind in a clearer light. In practice, however, the tints are often theoretically loaded, representing the fundamental assumptions of the sub-disciplines. Idiosyncratic concepts and techniques often prevent, rather than foster, a Jainian communication of viewpoints. Sub-disciplines are then analogous to the blind men facing the elephant: their grasp of reality remains distorted.

As a case of point, consider vision research. In the early days, it was commonly thought that knowledge about the external world influenced its perception. Theories of vision were "holistic," with perception resulting from interactions between low- and high-level visual processes (e.g., Helmholtz, 1856; Bruner & Goodman, 1947). However, significant advances arose from an analytic approach that isolates one specific visual process to test its performance envelope in tightly controlled conditions of experimentation. Nowadays, discoveries about the formation of face, object or scene categories, the attention to information in different recognition tasks, and the mechanics of recognition do not really inform research on the processes of low-level vision. Conversely, models of face, object and scene recognition and categorization are not

always firmly grounded on the established principles of early vision. However, to the extent that high- and low-level vision still examine different aspects of vision, they are not far from the blind men facing an elephant: They do not see it, and as importantly, they do not see each other!

Recent developments in vision research have witnessed the emergence of new concepts and techniques that could become a fruitful basis for communication between high- and low-level vision. These concern visual information and its selection to resolve high- and low-level visual tasks. We articulate these concepts and techniques into *RAP*, a new framework within which to formulate common issues. The main aim of this paper is to develop *RAP* and to re-establish a Jainian doctrine in vision.

RAP

To introduce *RAP*, we start with what is arguably the simplest recognition task: detecting whether or not a target object is present in the input. Suppose that the target is letter '𝔸' (standing for *A*vailable information). To detect it, assume that the observer uses a simple process of matching the incoming stimulus with only one template stored in memory. Assume further that the observer does not know '𝔸' and uses an '𝕣' (standing for *R*epresentation) template instead. With an appropriate matching criterion, the memorized template '𝕣' would successfully detect the presence of '𝔸' in the input.

In the process of matching '𝕣' to '𝔸', some information will be particularly potent. Note that we designed the example so that the letter '𝕡', the intersection of letters '𝔸' and '𝕣', corresponds to this potent information. Generalizing from this simple example, we articulate Represented (*R*), Available (*A*) and Potent (*P*) as follows:

$$R \quad A \approx P. \quad (1)$$

Equation (1) signifies that potent information P results from an interaction (represented by) of represented information R with available input information A. Potent information is an interesting, but little acknowledged construct in vision. P mediates visual categorization tasks; it is the information subset of A that can assign the unknown input to the category represented by R in memory. To illustrate, imagine that you need to categorize a face as smiling or not, as male or female, as young or old, or according to its identity. From a psychological viewpoint, potent information specifies the available information that must be particularly well attended to place a given face into this or that category (e.g., Gosselin, Archambault & Schyns, in press; Schyns & Oliva, 1999). This constraint on information from the categorization process could modify the tuning of vision to optimize the extraction of this particular information. For this reason we believe that $R \quad A \approx P$ (or RAP) is a useful bridge between the main information components of problems in vision: R specifies the high-level information constraints of a categorization task, A the low-level information available in the input, and P the necessary intermediate that the visual system must process to categorize the stimulus.

If the components of $R \quad A \approx P$ could be individually resolved, the equation would go a long way to bridge the gap between high- and low-level vision. At this stage, the reader might believe that the enterprise is no more than an utopia: representations are unobservable, and potent information varies across tasks and observers. To be successful, empirical handles are needed on the notoriously slippery components of RAP. The remaining of this paper will discuss recent developments that

provide the first tools to visualize aspects of *R* (*reverse correlation*) and *P* (*Bubbles*). We will introduce reverse correlation and *Bubbles,* in each case presenting illustrative examples of their application and generic algorithms for their operation. We then discuss the complementarity of the techniques and, finally, how they relate to the available information *A*. At the outset, it is worth pointing out that these tools are by no means a panacea. However, they nonetheless provide important empirical handles on important constructs of Cognitive Science (particularly *R* and *P*) that have so far proven elusive.

REVERSE CORRELATION

Wiener (1958) showed that noise could be used to analyze the behavior of a black box, even suggesting that the brain could be studied that way (Wiener, 1958, Lecture 8). His method is now known as reverse correlation. Later, Ahumada and Lovell (1971) refined the technique and made it suitable to applications in psychophysics.

To illustrate, imagine that we add uniform luminance noise to the letter 'A' to generate thousands of noisy stimuli. Adding 10% of such noise to our signal, for example, means replacing 10% of the pixels of 'A' by grey-level pixels sampled from a uniform distribution (with a mean equal to that of the original signal, see the top pictures of Figure 1). If we averaged the noisy letters, we would retrieve the original signal 'A', albeit with a lower contrast. Imagine an observer that has to decide whether or not a stimulus contains 'A'. He can make two possible responses ("yes, thesignal is present"; "no, the signal is absent") in each one of two possible conditions of stimulation (signal + noise, or noise alone). This leads to four possible classes of response (see the Table in Figure 1): a *Hit* (say "yes" when 'A' is present), a *False Alarm* (say "yes" when 'A' is absent), a *Miss* (say "no" when 'A' is present), or a *Correct Rejection* (say "no" when 'A' is absent).

Reverse correlation derives a *Classification Image* from the stimuli associated with the four classes of responses. Suppose we kept a record of all the stimuli that led to a hit and a false alarm. We add them together to obtain a *Yes Image* that depicts the information that elicited that response (see Figure 1). Likewise, we add the stimuli associated with a miss and a correct rejection to derive a *No Image* depicting the information that lead to that response (see Figure 1). The *Classification Image* is the difference between the *Yes Image* and the *No Image*. It depicts the representation that the observer used to detect '🅰' against noise (see the Figure 1).

Insert Figure 1 about here

The *Classification Image* in Figure 1 is an '🆁' (its representation), even though the original signal was only a noisy '🅰'. Where did the right leg of the '🆁' come from? The answer illustrates the power of reverse correlation: it reveals "hidden" information about a representation in memory when this representation is used to categorize the input. Specifically, when the memorized template '🆁' is matched against the input, input noise will sometimes be interpreted as the right leg of the memorized template (irrespective of whether or not the signal '🅰' is present). Across many trials, this interaction between the observer and the input reconstructs the missing leg of '🆁' from noise. The reconstruction happens because the representation '🆁' (not '🅰') drives the categorization process; it is a top-down reconstruction of represented structure. *False Alarms* and *Misses* are crucial for this reconstruction. Without them, *Hits* would only depict '🅰', and *Correct Rejections* would be a homogeneous image. In practice, noise is adjusted to a level that maintains about 25% categorization error (e.g., Beard & Ahumada, 1998).

Reverse correlation addresses the *R* of *R* $A \approx P$: It provides a behavioral handle on internal category representations. The technique has been successfully used in a number of domains ranging from features of auditory signals (Ahumada & Lovell, 1971; Pfafflin & Mathews, 1966; Ahumada & Lovell, 1971; Ahumada et al., 1975; Eggermont et al., 1983), electroretinograms (Sutter & Tran, 1992), visual simple response time (Simpson et al., 1998; Simpson et al., in press), single pulse detection (Thomas & Knoblanch, 1998), vernier acuity (Beard & Ahumada, 1998; Barth et al., 1999), stereopsis (Neri et al., 1999), letter discrimination (Gold, et al., 1999; Watson & Rosenholtz, 1997; Watson, 1998), single neuron's receptive field (e.g., Marmarelis & Naka, 1972; Jones & Palmer, 1987; DeAngelis, Ohzawa & Freeman, 1993; Eckhorn, Krause & Nelson, 1993; Emerson, Bergen & Adelson, 1992; McLean, Raab & Palmer, 1994; Ohzawa, DeAngelis, & Freeman, 1990), to scene discrimination (Ahumada & Beard, 1999; Abbey et al., 1999).

We will now illustrate the application of reverse correlation with two specific examples. The first one tackles the low-level vision problem of the shape (i.e., representation) of the receptive fields of LGN, simple cells and complex cells. The second example (from Gold, Murray, Bennett & Sekuler, 2000) addresses one problem in middle-level vision: the representation underlying the perception of illusory contours. Together, these examples illustrate the wide range of application of reverse correlation.

Reverse Correlation and the representation of receptive fields

In their seminal studies, Hubel and Wiesel (1959, 1962) examined the receptive fields of single cells in the visual pathway. The receptive field of a cell is the portion of the visual field to which the cell responds. Single cells in different parts of the visual stream respond optimally to different patterns of light impinging their receptive field. For example, the on-center/off-surround cell (LGN) (see leftmost image in Figure 2A; from a review article by Ohzawa, DeAngelis, & Freeman, 1995) fires maximally to the pattern of one spot of light hitting the central green area of its receptive field, and no

light in the surround red area (see also leftmost objects in Figures 2B and C, for other examples of receptive fields). Thus, a receptive field is a pattern against which incoming light is matched to drive the response of a cell.

Insert Figure 2 about here

Reverse correlation has recently been applied to precisely derive the shape of various receptive fields. For example, De Angelis, Ohzawa and Freeman (1993; see also Jones and Palmer, 1987) presented a continuous flow of pseudo-random stimuli (i.e., patterns of spots and bars) to the eyes of a cat, and continuously recorded the spikes of one neuron (e.g., in the LGN). Remember that reverse correlation reconstructs a representation from noise when this representation is used to match the input. DeAngelis et al. (1993) correlated the range of responses of a cell with the noisy inputs. In a nutshell, a high (vs. low) firing rate indicates a high (vs. low) match (i.e., correlation) between the noisy stimulus and the optimal pattern of light in the receptive field of the cell. The rightmost image in Figure 2A depicts the outcome of such a procedure. (The rightmost images in Figures 2B and C show the same for different types of cells.) It demonstrates not only that the receptive fields have their expected shapes, but also that the sensitivity of a given receptive field can be very precisely mapped.

Reverse correlation can thus be used to identify the representation underlying what is assumed to be the perception of the real contours (at the level of the neuron). We will see in the next section that it can also be applied to a problem at a higher level of visual organization: the perception of illusory contours—i.e., perceived contours that do not exist in the image.

Reverse correlation and illusory contours

Casual observers would experience no difficulty in classifying the leftmost and rightmost objects of the top row of Figure 3 as a thin square and a fat square,

respectively. For these judgements, they would use the available information *A* of real convex and concave contours. Perhaps more surprisingly, they can still resolve the task with only the sparse information of the second row. Here, observers perceive "illusory contours" appearing between the openings of the slightly rotated "Pacman"'s (Ringach & Shapley, 1996). Illusory contours are not present in the images; they are not in *A*, the available input information. However, are they present in *R*, the representation used to classify the squares as thin or fat?

Gold, Murray, Bennett and Sekuler (2000) used reverse correlation to address this question. Their observers saw fat and thin Kaniza squares to which noise was added (i.e., a proportion of pixels was replaced by random values taken from an independent Gaussian distribution). As explained earlier, the authors collected in separate bins the stimuli for which observers made *Hits*, *False Alarms*, *Misses* and *Correct Rejections*. After adding the *Hits* and *False Alarms* to obtain the *Yes Image*, and the *Misses* and *Correct Rejections* to obtain the *No Image*, they subtracted the *No Image* from the *Yes Image* to obtain a *Classification Image.*

Insert Figure 3 about here

Figure 3 summarizes the results of the experiment. The left (vs. right) column illustrates the different conditions of thin (vs. fat) signals. From top to bottom, the rows depict the conditions of REAL contours, ILLUSORY contours, OCCLUDED completion, TEXTURED OCCLUDED, and lastly FRAGMENTED. In each row, the central column represents the corresponding classification image (the red Pacmans are displayed for readability). Inspection reveals that all classification images but one (in the FRAGMENTED condition) demonstrate the use of represented vertical contours to

perform the thin vs. fat square task. However, these contours were only present in one condition of stimulation (REAL contours, see the top row)[1].

In sum, the examples have shown that reverse correlation can reveal the *R* of *R* *A* ≈ *P*. It does so via a reverse projection of a represented structure onto noise. In low-level vision, we have seen how this process could reconstruct the shape of receptive fields of LGN, simple and complex cells. Turning to middle-level vision, the technique revealed the representation underlying the perception of illusory contours. In both cases, the depiction of *R* comprised information that was not immediately visible in *A*. We now turn to another component of *RAP*, *P*, the potent information, and *Bubbles*, a new technique to visualize it.

BUBBLES

To introduce *Bubbles*, consider again the six blind men sparsely sampling an elephant. Assume now that they can communicate their respective viewpoint. Together, the blind man thinking that the elephant is a fan, and the blind man thinking that the elephant is a column might wrongly conclude that they are both facing a tree. However, the blind men thinking that they are respectively facing a snake, a rope and a wall could rightly infer that the object is an elephant. To categorize the elephant, the blind men would need to find the *potent* combinations of viewpoints. *Bubbles*' *modus operandi* is very much like blind men sparsely sampling an elephant: the technique samples a stimulus space while keeping track of the samples that lead to successful categorizations. Potent information is the expected outcome of *Bubbles*.

Going back to the letter example, suppose that '𝔄' is displayed behind an opaque mask punctured by a number of randomly located Gaussian holes (called "bubbles", see

[1] It is worth noting that, even though the contours in R between the pacmans in the ILLUSORY, OCCLUDED and TEXTURED OCCLUDED conditions were not present in A, there is still an overlap

Figure 4). The observer must decide whether or not the image behind the mask matches their represented 'ℝ'. In such conditions of sparse information, the bubbles may, or may not, reveal the potent information allowing the classification.

As in reverse correlation, we turn to the observer's responses to unravel the potent information. In *Correct Image*, we sum all the bubbles leading to *Hits* and *Correct Rejections*. In *Incorrect Image*, we sum all the bubbles leading to *Misses* and *False Alarms*. We then derive the *Proportion Image* by dividing the *Correct Image* by the sum of the *Correct* and the *Incorrect Images* (alternatively, you can use the arithmetically equivalent operation shown in Figure 4). The *Proportion Image* is an "attentional mask" that weights the importance of each region of the input space for the task at hand. In our letter example, the *Proportion Image* reveals the letter 'ℙ', the information of 'ℝ' that is potent to categorize it as 'ℝ' (see Figure 4). Note that 'ℙ' is the intersection of the represented information 'ℝ' with the available information 'ℝ'.

Insert Figure 4 about here

Categorization errors are important in *Bubbles*. If the observer performed the experiment without a single error, we would conclude that any random selection of bubbles contained sufficient information to classify the input. In contrast, categorization errors reveal the bubbles that did not contain potent information. In practice, the number of bubbles puncturing the opaque mask is adjusted to maintain an error rate of about 25% (see Gosselin & Schyns, 2000).

Bubbles for face identification

In high-level vision, Gosselin and Schyns (2000) used *Bubbles* to understand the potent information underlying face identification. To generate stimuli, we first

---

between A and R, i.e., the mouth of the pacmans.

decomposed the individual faces into 6 spatial frequency bandwidths of one octave each (see Figure 5a and b). The bandwidth containing the coarsest information was used as a background; it is not enough to recognize the faces. For each of the other bandwidths, we created an independent mask punctured with bubbles whose size was adjusted to the scale considered. The faces observers actually saw were reconstructed by adding the face information revealed at each scale by the bubbles. Observers' task was to identify the displayed face in a 10 Alternative-Forced-Choice paradigm (there were 10 different identities, 5 males and 5 females).

Insert Figure 5 about here

To derive P we computed a different *Proportion Image* per scale (see Figure 5c). For readability, we multiplied the scale information of Figure 5b with the potent masks of Figure 5c to obtain Figure 5d. At the finest scale, the eyes and a corner of the mouth stand out (see the leftmost picture in Figure 5d). At the next to finest scale the information comprises the eyes, the nose and the mouth. The next scale is closer to a configural representation of the face: Together, the eyes, the nose, the mouth and the chin form a meaningful recognition unit, but in isolation, none of the feature could identify the face (Gauthier & Tarr, 1997; Schyns & Oliva, 1999; Sergent, 1986; Tanaka & Sengco, 1997). At the next scale, the left side of the face silhouette is used–lighting was always at the right side of the faces and therefore their left sides were more shaded and more informative. The potent, or effective face is reconstructed by adding the potent information at each scale (see Figure 5 e).

In summary, *Bubbles* is a general technique that can assign the credit of a visual categorization task to its potent visual information. In $R \quad A \approx P$ potent information mediates between representations and available information. From a processing point

of view, the interaction between the human observer and randomly located bubbles can be depicted as a random search for potent information in the space of available information. With enough trials, a random search is exhaustive and all the search space is explored. It is worth pointing out that the search space can be abstract. It does not need to be tied to the image space (e.g., in the face example, one of the dimensions of the search is scale, but the search can be generalized to more abstract spaces, translation invariant, colored, and so forth).

RAPPING IT ALL UP

We have argued earlier that $R \quad A \approx P$ frames the information problem of visual categorization tasks. To be fruitful, *RAP* requires powerful handles on the nature of *R*, *P* and *A*. We have shown that reverse correlation provides a handle on memory representations and that *Bubbles* tackles potent information.

What is *A*, then? It depends on the context of interpretation. Keeping in line with an informational account of vision, *A* is the information formally useful to resolve a specific task. To illustrate, imagine an observer in a toy world who must identify 10 images in all points equal but for 100 varying grey-level pixels (see Box 1). Formally, only these 100 pixels form *A*, because they are useful to resolve the task. We can quantify *A* as 100 x 8 bits of information (if the grey scale had 256 levels)[2].

Suppose that our putative observer had a high-resolution memory. He could learn the 10 images at full resolution, using each one of the 100 informative grey-level pixels. The memory of each image would then require 100 x 8 bits of resources. To this ideal, omniscient (but unlikely) observer, memorized information would strictly equal

available information. Shortly put, $R \quad A \approx P$ would become $A \quad A \approx A$. Reverse correlation would reveal this equality, and so would *Bubbles*, because all of the informative pixels are equally potent to this ideal observer.

Human observers are not omniscient. They cannot represent the 10 images with all 100 grey-level pixels of formally available information. Instead, people represent visual events with sparse representations that comprise potent information for the task at hand. For example, one observer could notice that a particular subset of 10 pixels is potent to resolve the task, memorize the 10 images in terms of these 10 pixels, and dismiss entirely the remaining 90 available pixels. This reduction of information would transform $A \quad A \approx A$ into $P \quad A \approx P$, to reflect the fact that the memory representation of the images only comprise the 10 pixels of potent information for this observer (i.e., 10 x 8 bits of information per image).

Knowing A and P, we can now measure the efficiency (Kersten, 1990) of the human observer in contrast to the ideal observer. The efficiency is a measure of the information that the observer uses to resolve a task, given all the information available in this task. For the task of memorizing the 10 images, the efficiency is related to P/A—specifically, (10 images x 10 pixels x 8 bits) / (10 images x 100 pixels x 8 bits) = .1. Note that if the human observer was omniscient, its efficiency would become $P/A = A/A = 1$.

Potent information is the interaction between a representation and available information. Thus, the extraction of P from A is a prerequisite to a successful categorization of the incoming stimulus. However, if P provides sufficient conditions of

---

[2] To derive the structure of formally useful information in tasks of low-, mid- and high-level vision, we

categorization, it does not totally capture representations, and it will often be useful to differentiate the three components of $R \quad A \approx P$. The human observer can for example notice that, in one particular face, some pixels falling outside the available information frame have a certain grey value. These pixels would neither be part of A (because they are constant across images and thus not formally useful), nor part of P (because useless pixels cannot be potent). However, the human observer might nevertheless represent this particular face as such, for example because this information contrasts with other categories he has represented in memory. In this case, R would be different from P and A, hence $R \quad A \approx P$.

Situations of recognition are similar to this latter case. In a given laboratory task, A is the formally measurable information, and P the potent subset of A. As explained earlier, these two information sources are in principle measurable, comparable (e.g., in a measure of efficiency), and derivable from the ideal observer and *Bubbles*. What about R? Reverse correlation can reveal the aspects of the representation that are used in the considered recognition task. However, there are aspects of R that reverse correlation would not reveal. Unknown to the experimenter, the observer can represent the experimental objects in a broader context than that of the specific experimental task. The observer knows other stimulus categories and can represent properties of the stimuli which, very much like the black pixels above, are not strictly useful to the task at hand (memorize the 10 images) but do distinguish this class of image from other category representations in memory. Obviously, if the observer does resolve the task, he has represented P. But the point is that R might be more than P, and this is a general

---

refer the reader to the ideal observer approach (Kersten, 1990).

problem for studying representations: reverse correlation can only reveal the components of *R* that are used in a specific experimental task, not those of the broader conceptual system in which this category is inserted. For this reason, we predict that reverse correlation and *Bubbles* will in many instances reveal the same information: *P*.

The discussion has so far focused on the informational interpretation of *R* *A* ≈ *P*, and we have briefly discussed how available information and its use could be quantified. We now turn to another possible reading of *R* *A* ≈ *P*: *Representation* *Available information* ≈ *Perception*. To this end, we must change the meaning of *R* *A* ≈ *P* from a strict informational account to bring it closer to the phenomenology of perception. Instead of treating *A* as the formally useful information, it becomes *all* the visual information that is derivable from the retinal projection of the distal stimulus.

To bridge between the informational and perceptual interpretations of *R* *A* ≈ *P*, consider the special role of potent information. We already argued that potent information mediates a categorization problem. Thus, attention to and perception of *P* is a prerequisite to a successful recognition of the stimulus. Consequently, the perception of this stimulus will comprise information on a continuum somewhere between the sparse *P* and the complete *A*. As Hochberg (1982) eloquently put it: "… unlike objects themselves, our perception of objects are not everywhere dense. " To the extent that potent information itself depends on the task being resolved, it depends on the representation that is used. For a given stimulus, one might find that the potent

information changes dramatically across categorization tasks. $R \quad A \approx P$ predicts that the perception of the stimulus might change accordingly.

CONCLUSIONS

This paper has developed *RAP*, a framework to characterize problems in vision. We have seen how reverse correlation provides a behavioral handle on representations, *Bubbles* tackles potent information, and suggested that the ideal observer could characterize the available information. Our own technique emphasizes the role of potent information, and we would contend that the failure to acknowledge the role of potent information is one of the reasons underlying the schism between the domains of high- and low-level vision.

REFERENCES

Abbey, C. K., Eckstein, M. P., and Bochurd, F. O. (1999). Estimation of human-observer templates in two-alternative forced-choice experiments. In E. A. Krupinski (Ed.), *Proceedings of the Society of Photo-optical Instrumentation Engineers*, pp. 284-295, San Diego: SPIE.

Ahumada, A. J. and Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America*, **49**, 1751-1756.

Ahumada, A. J., Marken, R. and Sandusky, A. (1975). Time and frequency analyses of auditory signal detection. *Journal of the Acoustical Society of America*, **57**, 385-390.

Backstein, K. (1992). *The Blind Men and the Elephant*. New York: Scholastic.

Barth, E., Beard, B. L. and Ahumada, A. J. (1999). Nonlinear features in vernier acuity. In B. E. Rogowitz and T. N. Pappas (Eds.), *Human Vision and Electronic Imaging IV*, *SPIE Proceedings*, **3644**, paper 8.

Beard, B. L. and Ahumada, A. J. (1998). A technique to extract the relevant features for visual tasks. In B. E. Rogowitz and T. N. Pappas (Eds.), *Human Vision and Electronic Imaging III, SPIE Proceedings*, **3299**, 79-85.

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

DeAngelis, G. C., Ohzawa, I. and Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology*, **69**, 1091-1117.

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

Duda, R. O and Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons.

REFERENCES

Abbey, C. K., Eckstein, M. P., and Bochurd, F. O. (1999). Estimation of human-observer templates in two-alternative forced-choice experiments. In E. A. Krupinski (Ed.), *Proceedings of the Society of Photo-optical Instrumentation Engineers*, pp. 284-295, San Diego: SPIE.

Ahumada, A. J. and Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America*, **49**, 1751-1756.

Ahumada, A. J., Marken, R. and Sandusky, A. (1975). Time and frequency analyses of auditory signal detection. *Journal of the Acoustical Society of America*, **57**, 385-390.

Backstein, K. (1992). *The Blind Men and the Elephant*. New York: Scholastic.

Barth, E., Beard, B. L. and Ahumada, A. J. (1999). Nonlinear features in vernier acuity. In B. E. Rogowitz and T. N. Pappas (Eds.), *Human Vision and Electronic Imaging IV*, *SPIE Proceedings*, **3644**, paper 8.

Beard, B. L. and Ahumada, A. J. (1998). A technique to extract the relevant features for visual tasks. In B. E. Rogowitz and T. N. Pappas (Eds.), *Human Vision and Electronic Imaging III, SPIE Proceedings*, **3299**, 79-85.

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

DeAngelis, G. C., Ohzawa, I. and Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology*, **69**, 1091-1117.

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

Duda, R. O and Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons.

Eckhorn, R., Krause, F. and Nelson, J. I. (1993). The RF-cinematogram. A cross-correlation technique for mapping several visual receptive fields at once. *Biological Cybernetics*, **69**, 37-55

Eggermont, J. J., Johannesma, P. I. M. and Aertsen, A. M. H. J. (1983). Reverse-correlation methods in auditory research. *Quaterly Reviews of Biophysics*, **16**, 341-414.

Emerson, R. C., Bergen, J. R. and Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research*, **32**, 203-218.

Gauthier, I., & Tarr, M. J. (1997). Becoming a 'Greeble' expert: exploring the face recognition mechanism. *Vision Research*, **37**, 1673-1682.

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

Gosselin, F. and Schyns, P. G. (2000). *Bubbles: a technique to reveal the use of information in recognition tasks.* Manuscript submitted for publication.

Gosselin, F., Archambault, A. and Schyns, P. G. (in press). Interactions between taxonomic knowledge, categorization, and perception. In M. Ramscar, U. Hahn, E. Cambouropolos, & H. Pain (Eds.) *Similarity and categorization*. Cambridge: Cambridge University Press.

Helmholtz, H. (1856-1866). *Handbuch der Physiologischen Optik*. Leipzig: Voss.

Hochberg, J. (1982). How big is a stimulus. In J. Beck (Ed.) *Organization and Representation in Perception*, 191-217. Lawrence Erlbaum, NJ.

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology*, **148**, 574-591.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology*, **160**, 106-154.

Jones, J. P. and Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, **58**, 1187-1211.

Kersten, D. (1990). Statistical limits to image understanding. In Colin Blakemore (Ed.) *Vision: Coding Efficiency*. 32-44. Cambridge: Cambridge University Press.

Marmarelis, P. Z. and Naka, K.-I. (1972). White-noise analysis of a neuron chain: An application of the Wiener theory. *Science*, **175**, 1276-1278.

McLean, J., Raab, S. and Palmer, L. A. (1994). Contribution of linear mechanisms to the specification of local motion by simple cells in areas 17 and 18 of the cat. *Visual Neuroscience*, **11**, 271-294.

Neri, P., Parker, A. J. and Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature*, **401** (6754), 695-698.

Oshawa, I., DeAngelis, G. C. and Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science*, **249** (4972), 1037-1041.

Oshawa, I., DeAngelis, G. C. and Freeman, R. D. (1995). Receptive-field dynamics in the central visual pathways. *Trends in Neuroscience*, **18**, 451-458.

Pfafflin, S. M and Mathews, M. V. (1966). Detection of auditory signals in reproducible noise. *Journal of the Acoustical Society of America*, **39**, 340-345.

Ringach, D. L. and Shapley, R. (1996). Spatial and temporal properties of illusory contours and amodal boundary completion. *Vision Research*, **36**, 3037-50

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

Sergent, J. (1986). Microgenesis of face perception. In H.D. Ellis, M.A. Jeeves, F. Newcombe, & A.M. Young (Eds.), *Aspects of face processing.* Dordrecht: Martinus Nijhoff.

Simpson, W. A., Braun, J., Bargen, C. and Newman, A. (in press). Identification of the eye-brain-hand system with point processes: A new approach to simple reaction time. *Journal of Experimental Psychology: Human Perception and Performance*.

Sutter, E. E. and Tran, D. (1992). The field topography of ERG components in man–I: The photopic luminance response. *Vision Research*, **32**, 433-446.

Tanaka, J. and Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition.*

Thomas, J. P and Knoblanch, K. (1998). What do viewers look for when detecting a luminance pulse? *Investigative Opthalmology and Visual Science*, **39**, S404.

Watson, A. B. (1998). Multi-category classification: template models and classification images. *Investigative Opthalmology and Visual Science*, **39**, S912.

Watson, A. B. and Rosenholtz, R. (1997). A Rorschach test for visual classification strategies. *Investigative Opthalmology and Visual Science*, **38**, S1.

Wiener, N. (1958). *Nonlinear Problems in Random Theory*. New York: Wiley.

FIGURE CAPTIONS

Figure 1. (a) Three signal + noise (70% of uniform luminance noise) stimuli and three noise-only stimuli. (b) The four classes of possible response, and the steps involved in the computation of the **ClassificationImage**. Observer's responses were obtained by computer simulation. For each noisy stimulus, the model computed the Euclidean distance to the 'R' template, and decided to respond "yes, present" (when the stimulus was closer to the mean of the signal + noise distribution) or "no, absent" (when the stimulus was closer to the mean of the noise-only distribution). Noise level was adjusted to 95% maintain performance at 77% correct.

Figure 2. Spatial Receptive Field (RF) structure of the major classes of neurons in the geniculostriate pathway from Ohzawa, DeAngelis and Freeman (1995). (A) Schematic and experimental profiles of the RF of an ON-center neuron from the LGN of a cat. In the traditional depiction (left), the RF has a central "ON" region (+) which is responsive to the onset of a bright stimulus, and a surrounding "OFF" region (-) which is responsive to the onset of a dark stimulus (or the offset of a bright stimulus). On the right is shown a 2-D spatial (X-Y) RF profile for an ON-center X-cell, as measured using a reverse correlation technique. Regions of visual space that are responsive to bright spots are delimited by solid contour lines; regions responsive to dark spots are represented by dashed contours. Darkness of shading is proportional to response strength. A center-surround structure is clearly seen in this profile, although the surround is fairly weak. Similar data have been presented elsewhere for retinal ganglion58 and LGN 9,36,59 cells. (B) Depicted schematically on the left, the RF of a simple cell exhibits an alternating arrangement of elongated sub-regions that are responsive to either bright (+) or dark (-) stimuli. A measured RF profile for a simple cell from cat striate cortex (area 17) is shown on the right as a contour map (conventions as in A). Similar data have been presented elsewhere. (C) Spatial RF structure of a complex cell. In the traditional schematic illustration shown on the left, pluses and minuses are shown throughout the field,

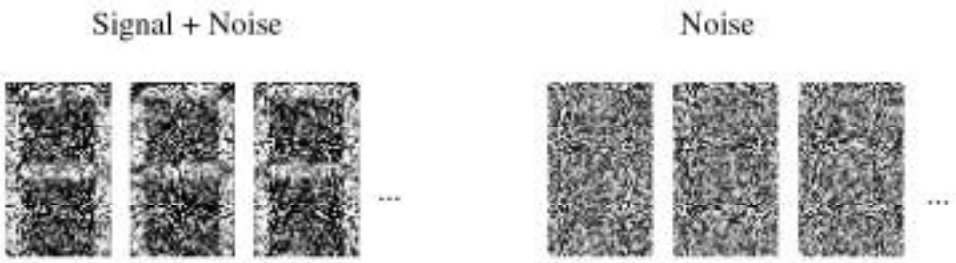
indicating that the cell responds to both bright and dark stimuli at each position. Panels on the right show the RF pro-file of an area 17 complex cell, as measured using reverse correlation (see also Refs. 17,55). Because regions responsive to bright and dark stimuli overlap, separate profiles are shown for bright and dark stimuli.

Figure 3. Illustration of stimuli and results from Gold et al. (2000), mapping the behavioral receptive field for visually completed contours. Each row corresponds to a different condition. The left and right columns show thin and fat stimuli, respectively. In the experiment, each corner Pacman was rotated by ±1.75°. The Pacmans of the stimuli shown in the figure have been rotated by ±3.5° for clarity. The middle column shows smoothed average classification images, combining data from three observers. In other words, it shows the locations of the stimuli observers used in discriminating thin and fat; black and white pixels indicate the most significant locations. Red inducers have been superimposed on each classification image. Note that for all conditions except the FRAGMENTED one, observers use information in essentially the same locations, although there is no physical contour located there in the case of ILLUSORY, OCCLUDED, and TEXTURED OCCLUDED objects. This is the first direct evidence for a unique processing unit for illusory and real contours.

Figure 4. (a) Three signal-revealed-by-bubbles (50 Gaussian bubbles with a 2 pixels standard deviation) stimuli and as many bubbles alone stimuli. (b) The four classes of response, and the steps involved in the computation of the **ProportionImage**. The data was obtained by computer simulation. For each bubbled image, the model computed the Euclidean distance to an 'R' template (with bubbles), and responded "yes, present" when the stimulus was closer to the mean of distances of signal-revealed-by-bubbles, and "no, absent", when the stimulus was closer to the mean of distances of bubbles-only. The number of Gaussian bubbles with a 1 pixel standard deviation was adjusted to 3 to keep the performance of the rapper at 76% correct responses.

Figure 5. This figure illustrates the outcome of *Bubbles* in Experiment 2 of Gosselin and Schyns (2000). Participants learned 10 identities. They then attempted to identify the faces revealed by bubbles. The number of bubbles was adjusted to maintain the performance constant at 75% correct responses. Pictures in (b) represent five independent scales of (a) at 5.62, 11.25, 22.5, 45, 90 cycles per face, from coarse to fine. (c) represent the statistically significant potent regions at each spatial scale. (d) is the product of (b) by (c). The bottom picture is the potent stimulus: a depiction of the information used to identify faces in Experiment 2.

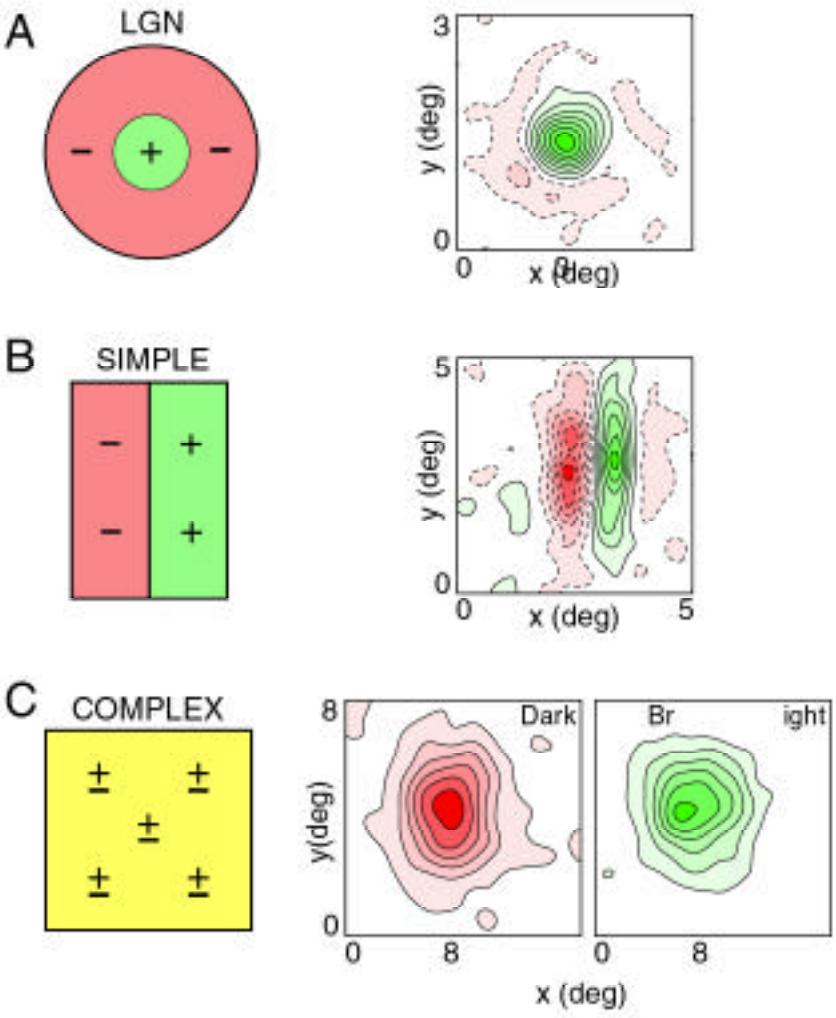
a.

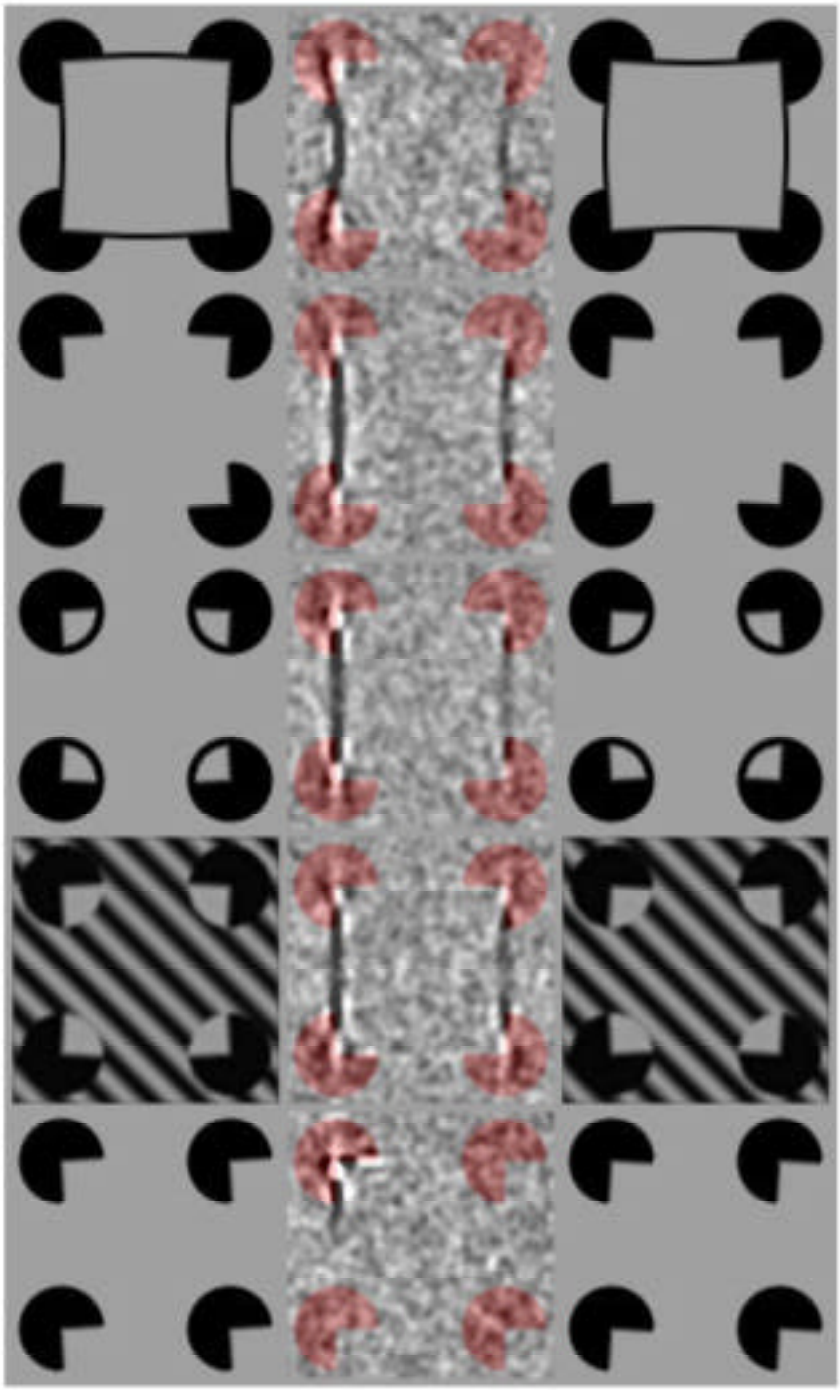


b.

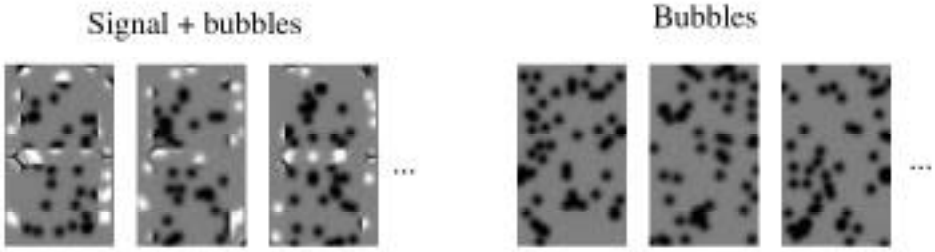
	Incorrect	Correct
Yes	**FalseAlarms** (573) +	**Hits** (1909) = **YesImage** -
No	**Misses** (558) +	**CorrectRejections** (1960) = **NoImage** =

ClassificationImage


A
LGN
−
+
−
3
y (deg)
0
0
x (deg)
B
SIMPLE
−
+
−
+
5
y (deg)
0
0
x (deg)
5
C
COMPLEX
±
±
±
±
±
8
y(deg)
0
Dark
Br
ight
0
8
0
8
x (deg)

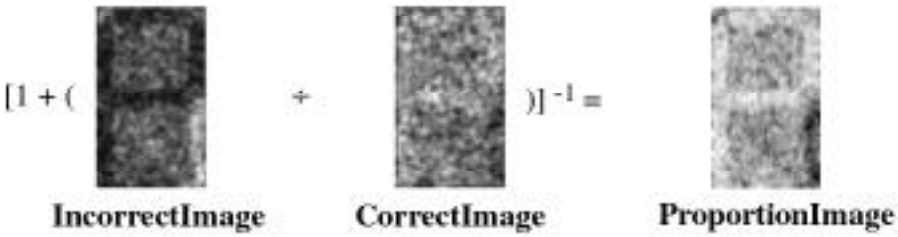


a.



b.

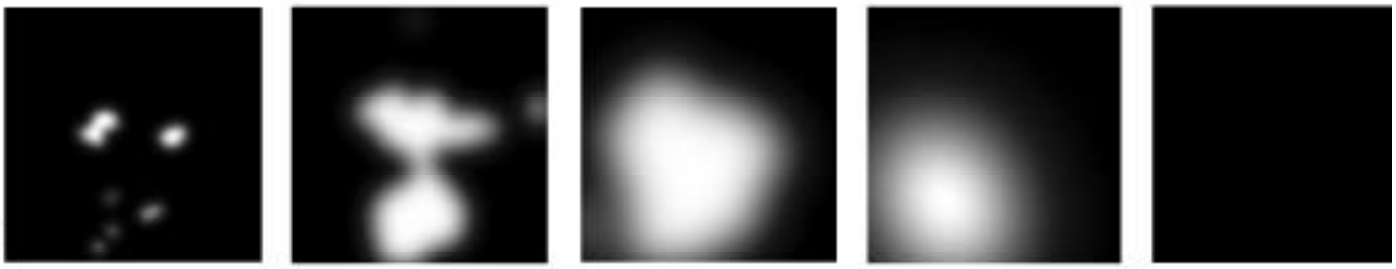
	Incorrect	Correct
Yes	**FalseAlarms** (547) +	**Hits** (1889) +
No	**Misses** (632)	**CorrectRejections** (1932)



a.

b.

c.



d.

e.

$\mathcal{P}$:

3. Superstitious Observer: $\mathcal{R} \otimes \mathcal{A} \approx \mathcal{P}$

$\mathcal{R}$:

$\mathcal{A}$:

$\mathcal{P}$: