RAP: a new framework for visual categorization

Frédéric Gosselin and Philippe G. Schyns

Cognitive science might almost be defined as several disciplines communicating their different perspectives on the mind, the common object of study. However, domain-specific concepts and techniques can prevent, rather than foster, a communication of viewpoints. In this article, we develop a new framework for visual categorization in which the interaction between Represented ($\mathcal{R}$) information and Available ($\mathcal{A}$) information determine the Potent ($\mathcal{P}$) information (symbolically, $\mathcal{R}\otimes\mathcal{A}\approx\mathcal{P}$). We argue and illustrate that this framework helps to establish a common language to articulate issues common to low-, mid-, and high-level vision. More importantly, we present new techniques 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 [1]. 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 touches a leg and infers that it is a column; a fifth the tail, concluding that it is a frayed bit of rope; and the sixth grasps one of the ears and believes it to be 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 incomplete and potentially 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 [2,3]. 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 mechanisms 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 run the risk of being like the blind men facing an elephant.

'...models of face, object and scene recognition and categorization are not always firmly grounded on the established principles of early vision.'

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 developments 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 advocate the need of a Jainian doctrine in vision science.

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 this observer does not know 'A' but instead uses an approximation of 'A' that happens to have the shape of an 'R' (standing for *R*epresentation). This memorized 'R' could successfully detect 'A' in the input, with an appropriate matching criterion.

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' (standing for *P*otent), the intersection of letters 'A' and 'R', corresponds to this information. Generalizing from this simple example, we articulate Represented ($\mathcal{R}$), Available ($\mathcal{A}$) and Potent ($\mathcal{P}$) as follows:

$$\mathcal{R}\otimes\mathcal{A}\approx\mathcal{P} \quad (1)$$

Equation (1) signifies that potent information, $\mathcal{P}$, results from an interaction (represented by $\otimes$) between represented information, $\mathcal{R}$, and available

Frédéric Gosselin
Dépt de Psychologie, Université de Montréal, C. P. 6128, Succ. Centre-ville, Montréal QC, Canada H3C 3J7.
e-mail: frederic.gosselin@umontreal.ca

Philippe G. Schyns
Dept of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow, UK G12 8QB.
e-mail: philippe@psy.gla.ac.uk

http://tics.trends.com 1364-6613/02/$ – see front matter PII: S1364-6613(00)01838-6

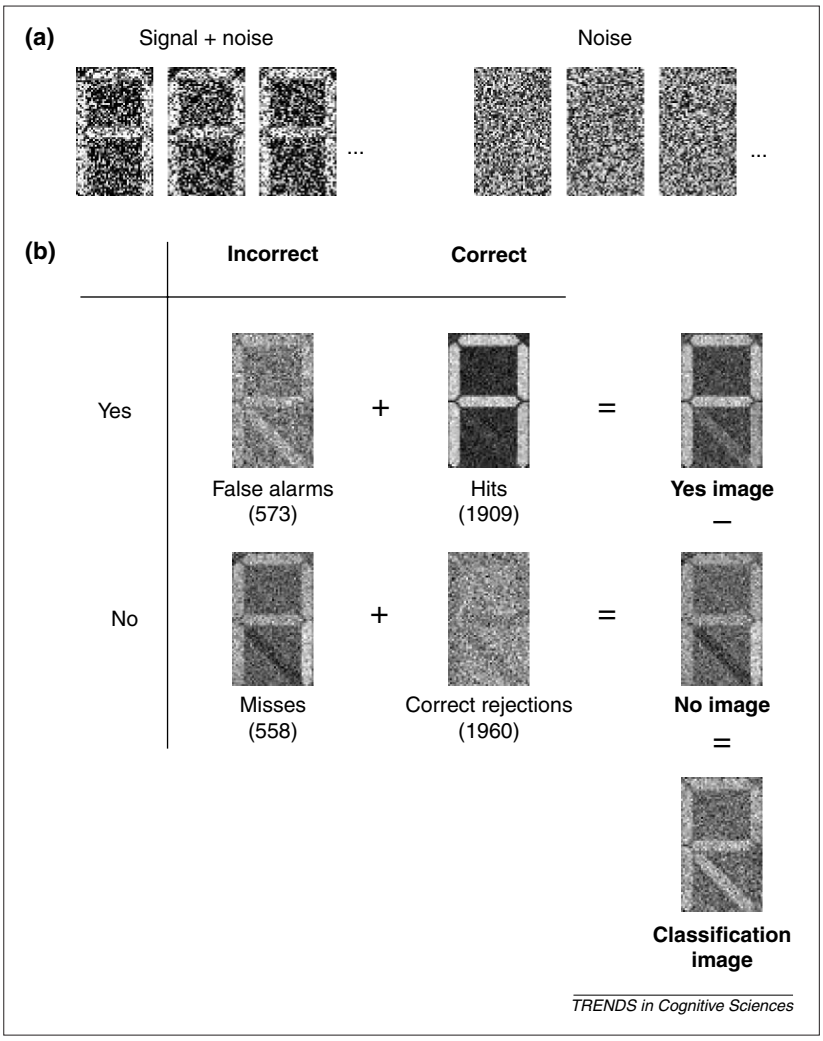

Fig. 1. (a) Three signal + (Gaussian) noise stimuli and three noise-only stimuli. (b) The four classes of possible response, and the steps involved in the computation of the *Classification Image* (see text). Computer simulations produced the observer's responses. For each noisy stimulus, the model computed the city-block distance to the 'R' template that approximates the 'A' signal. It positively matched 'R' with 'A' (a '*yes*' detection response) whenever the stimulus was closer to the mean of the signal + noise distribution. By contrast, when the stimulus was closer to the mean of the noise-only distribution, the model produced a '*no*' rejection response. Noise level was adjusted to maintain performance at ~75% correct.

input information, $\mathcal{A}$. Potent information is an interesting, but little acknowledged construct in vision. $\mathcal{P}$ mediates visual categorization tasks; it is the information subset of $\mathcal{A}$ that can assign the unknown input to the category represented by $\mathcal{R}$ 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 standpoint, potent information specifies the visual information that must be particularly well attended to in order to place a given face in this or that category [4,5]. This constraint on the categorization process on the information needed could modify the tuning of vision to optimize the extraction of this information in the input. For this reason we believe that $\mathcal{R} \otimes \mathcal{A} \approx \mathcal{P}$ (or $\mathcal{RAP}$) is a useful bridge between the main information components of problems in vision: $\mathcal{R}$ specifies the high-level information requirements of a categorization task, $\mathcal{A}$ the information available in the input to resolve this task, and $\mathcal{P}$ the subset of $\mathcal{A}$ that the human visual system must process (i.e. attend to, extract and recode) to categorize the stimulus.

If each component of $\mathcal{R} \otimes \mathcal{A} \approx \mathcal{P}$ 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 a utopia: representations are unobservable, and potent information should vary across tasks and observers. To be successful, empirical handles are needed on the notoriously slippery components of $\mathcal{RAP}$. The remainder of this paper discusses recent developments that provide the first tools to visualize aspects of $\mathcal{R}$ ('*reverse correlation*') and $\mathcal{P}$ ('*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 that have so far proven elusive.

Reverse correlation

Wiener showed that noise could be used to analyze the behavior of a black box, even suggesting that the brain could be studied that way [6]. His method is now known as reverse correlation. Later, Ahumada and Lovell [7] modified the technique and made it suitable to applications in psychophysics.

To illustrate, imagine that we add Gaussian white contrast noise to the letter 'A' to generate thousands of noisy stimuli. (Adding such noise when the Gaussian distribution has a mean of zero and standard deviation of 0.1, for example, means that 95% of the pixels of the 'A' will be less than 0.165 away from the original contrast, see Fig. 1a.) If we averaged the noisy letters, we would retrieve the original signal 'A'. Imagine an observer that has to decide whether or not a stimulus contains 'A'. He can make two possible responses ('*yes, the signal is present*' or '*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 Fig. 1b): 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 Fig. 1b).

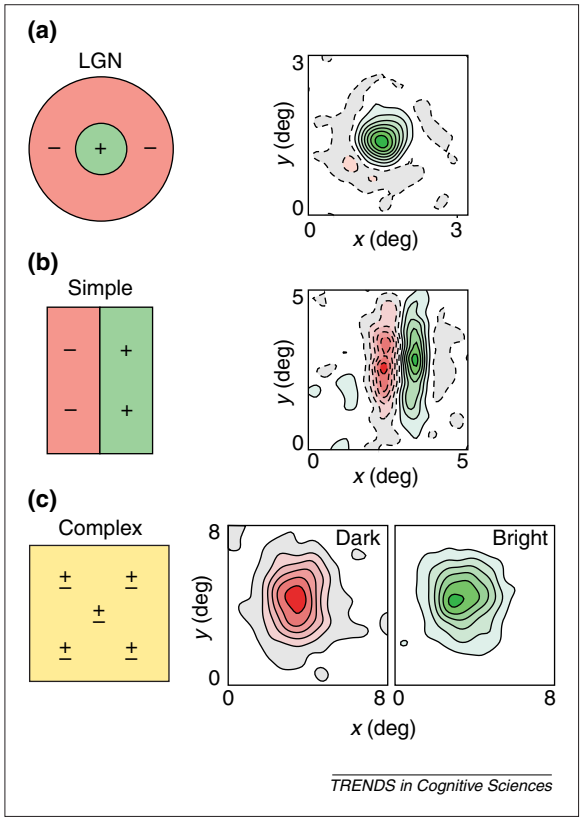

Fig. 2. Spatial receptive field (RF) structure of the major classes of neurons in the geniculostriate pathway. (a) Schematic and experimental profiles of the RF of an ON-center neuron from the LGN of a cat. Left, in the traditional depiction, the RF has a central 'ON' region (+), responsive to the onset of a bright stimulus, and a surrounding 'OFF' region (–), responsive to the onset of a dark stimulus (or the offset of a bright stimulus). Right, a 2-D spatial (*x*,*y*) RF profile for an ON-center LGN 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. (b) Left, the spatial RF of a simple cell in visual cortex V1 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). (c) Left, spatial RF structure of a complex cell. Pluses and minuses are shown throughout the field, indicating that the cell responds to both bright and dark stimuli at every position. Right, the RF profile of an area 17 complex cell, as measured using reverse correlation. Because regions responsive to bright and dark stimuli overlap, separate profiles are shown for bright and dark stimuli. (Adapted from DeAngelis, Ohzawa and Freeman [28].)

Likewise, we add the stimuli associated with a miss and a correct rejection to derive a **No Image** depicting the information that led to that response. The **Classification Image** is the difference between the **Yes Image** and the **No Image**. It depicts the representation that the observer used to detect 'A' against noise.

The **Classification Image** in Fig. 1 depicts an 'R' (for *R*epresentation), even though the original signal was only a noisy 'A' (for *A*vailable). Where did the right leg of the 'R' come from? The answer illustrates the power of reverse correlation: it can reveal 'hidden' information about a representation in memory when this representation is used to categorize the input. Specifically, when the observer matches the memorized template 'R' against the input, noise in the input is sometimes interpreted as the right leg of the memorized template (irrespective of whether or not the signal 'A' is present). Across many trials, this interaction between the observer and the input reconstructs the missing leg of 'R' from noise. The reconstruction happens because the representation 'R' (not 'A') drives the categorization process; it is a top-down reconstruction of represented memory structures. False Alarms and Misses are crucial for this reconstruction. Without them, Hits would only depict 'A', and Correct Rejections would be a homogeneous image. In practice, noise is adjusted to a level that maintains about 25% categorization error [8].

'Reverse correlation... can reveal "hidden" information about a representation in memory when this is used to categorize the input.'

Reverse correlation addresses the $\mathcal{R}$ of $\mathcal{R} \otimes \mathcal{A} \approx \mathcal{P}$: 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 [7,9–11], electroretinograms [12], visual simple response time [13], single pulse detection [14], vernier acuity [8,15], stereopsis [16], letter discrimination [17–19], single neuron's receptive field [20–27], see [28] for a review, modal and amodal completion [29], face recognition [19,30], scene discrimination [31], to attention [32].

We will now illustrate the application of reverse correlation with three 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 [29] addresses one problem in middle-level vision: the representation underlying the perception of illusory contours. The third example from [19] turns to high-level vision to depict the properties of object representations. Together, these examples illustrate the wide application of reverse correlation.

Reverse correlation and the representation of receptive fields

In their seminal studies, Hubel and Wiesel examined the receptive fields of single cells in the visual pathway [33,34]. 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 Fig. 2a, left) fires maximally to the pattern of one spot of

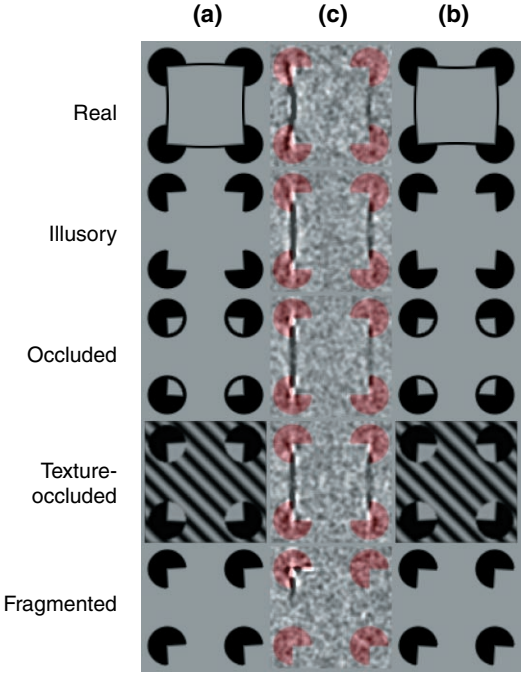

Fig. 3. Illustration of stimuli and results from Ref. [29], mapping the behavioral receptive field for visually completed contours. Each row corresponds to a different condition. Columns (a) and (b) show thin and fat stimuli, respectively. In the experiment, each corner Pacman was rotated by ±1.75°. The Pacmen of the stimuli shown in the figure have been rotated by ±3.5° for clarity. Column (c) 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 for visibility.) 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.

light hitting the central green area of its receptive field, and no light in the surround red area (see also Fig. 2b and c, left images, 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.

Reverse correlation has recently been applied to precisely derive the shape of various receptive fields. For example, DeAngelis, Ohzawa and Freeman [22] 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) (see also [21]). Remember that reverse correlation reconstructs a representation from noise when this representation is used to match the input. DeAngelis *et al.* correlated the range of responses of a cell with the noisy inputs [22]. 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 Fig. 2a depicts the outcome of such a procedure. (The right images in Fig. 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 the technique has also been applied to a problem at a higher level of visual organization: the perception of illusory contours – that is, perceived contours that do not physically 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 Fig. 3 as a thin square and a fat square, respectively. For these judgments, 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 'Pacmen' [35]. 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 used reverse correlation to address this question [29]. Their observers saw fat and thin Kaniza squares embedded in Gaussian white noise. 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**.

Figure 3 summarizes the results of the experiment. The left and right columns illustrates the different conditions of 'thin' and 'fat' signals, respectively. From top to bottom, the rows depict the conditions of real contours, illusory contours, occluded completion, textured-occluded completion, and fragmented images. In each row, the central column represents the corresponding classification image (the red Pacmans are displayed for visibility). Inspection reveals that all classification images but one (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, top row).

Reverse correlation and object representations

We applied reverse correlation to the higher-level problem of depicting the properties of internal object representations in memory [19]. In Experiment 1, an observer was instructed to detect the presence of the

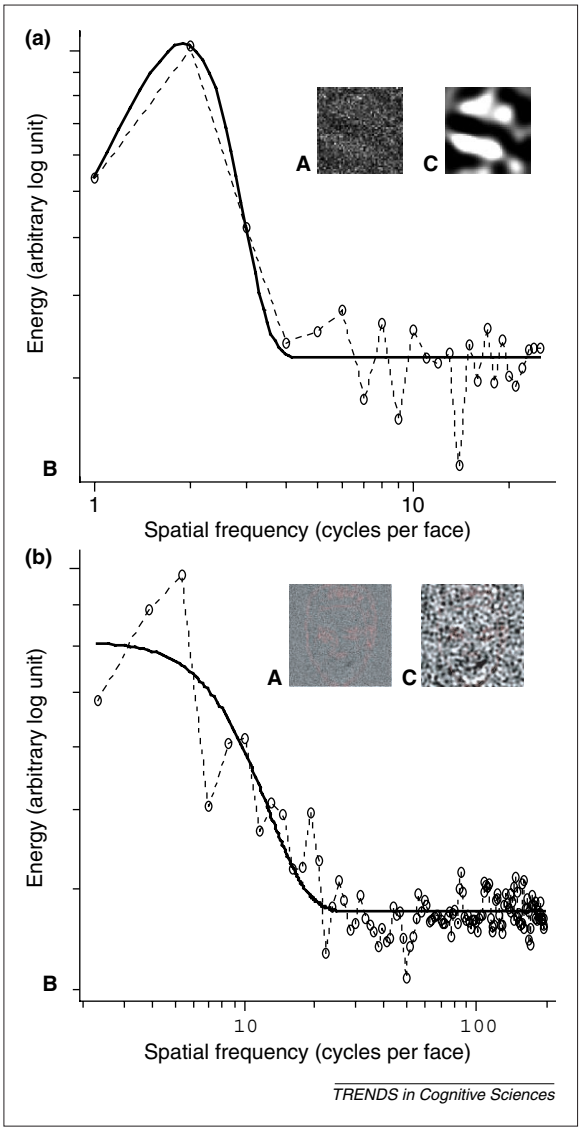

Fig. 4. (a) In the letter experiment of Ref. [19], observers only saw 50 × 50 static-bit, white-noise images. A, raw classification image; B, distribution of the average squared amplitude energy for different spatial frequencies of the classification image (expected energy = constant); C, classification image filtered with a low-pass (Butterworth) filter with a cut-off at 3 cycles per letter. A black 'S' on a white background is revealed. (b) In the smiling face experiment, the white noise comprised 27.5% of the black pixels of the contours of a mouthless face (indicated with a red marker in A and C) randomly sampled. A, raw classification image; B, distribution of the average squared amplitude energy for different spatial frequencies of the classification image (expected energy = constant); C, classification image filtered with a low-pass (Butterworth) filter with a cut-off at 30.52 cycles per face. A smile showing the teeth is revealed.

letter 'S' on half of the trials. In Experiment 2, an observer had to discriminate a smiling, from a non-smiling face. However, and this is crucial to the design of these experiments, observers were only stimulated with white noise fields. That is, no signal (i.e. a 'S' letter or a smiling mouth) was ever added to the fields of white noise. However, observers firmly believed that they perceived an 'S' letter in Experiment 1, and a smiling face in Experiment 2. As a white noise sequence does not represent any coherent structures in the image plane, these superstitious perceptions must arise from the observer's share. Using reverse correlation, we were able to depict the observer's internal representation of an 'S' and a smiling mouth (see Fig. 4). Interestingly, the spectral composition of the classification images corresponded in both cases to the most efficient bandwidth for the recognition of natural letters and face expressions [36–38].

Note that this application (and that of Ref. [22]) captures the essence of the reverse correlation technique: classified noise enables a reconstruction of hidden structures. This is closer in spirit to Wiener's original proposal and presents the main advantage that it cannot introduce any bottom-up (signal) bias in the classification responses. The main disadvantage, however, is that classic psychophysical methods cannot be applied (e.g. d' computations).

In summary, the examples have shown that reverse correlation can reveal the $\mathcal{R}$ of $\mathcal{R} \otimes \mathcal{A} \approx \mathcal{P}$. 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 high-level vision, it revealed the properties of an 'S' and a smiling mouth represented in memory. In all cases, the depiction of $\mathcal{R}$ comprised information that was not immediately visible in $\mathcal{A}$. We now turn to another component of $\mathcal{RAP} - \mathcal{P}$, the potent information, and *Bubbles*, a new technique to visualize it.

Bubbles

To introduce *Bubbles*, consider again the six blind men sparsely sampling information from 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 similar to this: 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 Fig. 5a). 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.

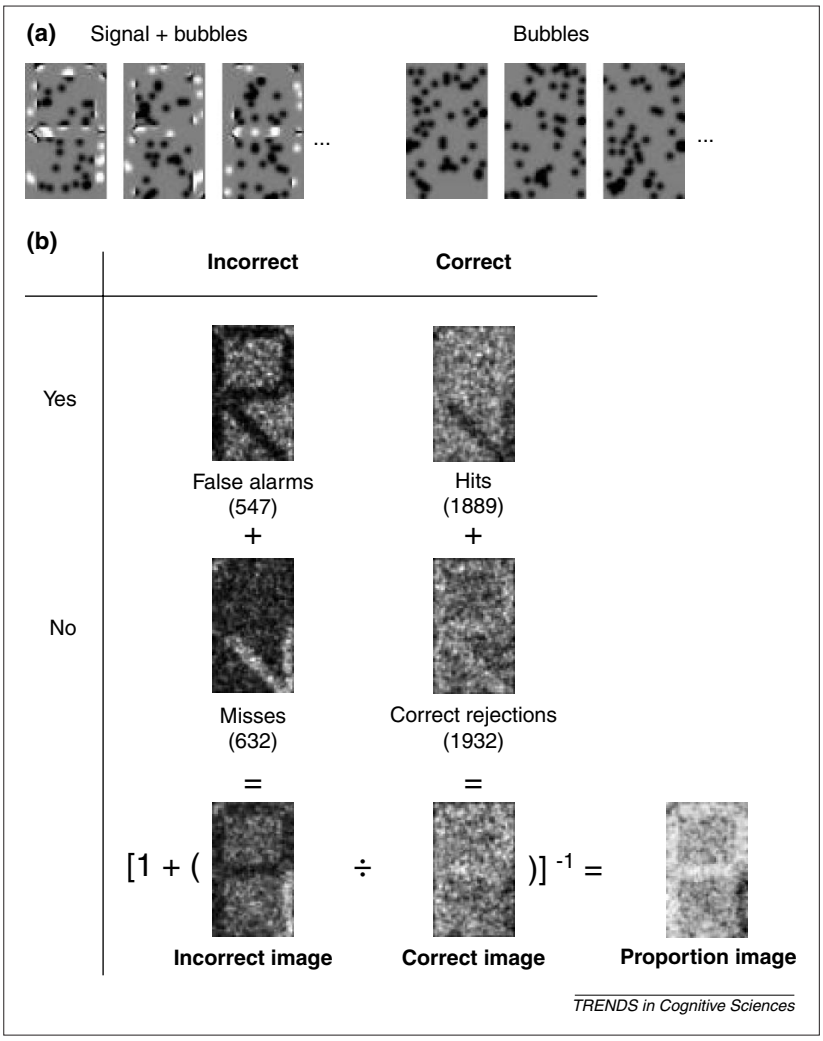

Fig. 5. (a) Three stimuli of signals revealed by bubbles (50 Gaussian bubbles with a 2 pixels standard deviation) and, on the right, bubbles alone stimuli. (b) The four classes of response, and the steps involved in the computation of the **Classification Image**. The data were obtained by computer simulation. For each bubbled image, the model computed the city-block distance to an 'R' template (with bubbles) that approximates the 'A' signal, and responded '*Yes, present*' when the stimulus was closer to the mean of the distances of the signal-revealed-by-bubbles, and '*No, absent*', when the stimulus was closer to the mean of the distances of the bubbles only. The number of Gaussian bubbles with a standard deviation of one pixel was adjusted to three to keep the performance of the observer at ~75% correct responses.

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 (see Fig. 5b). 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** (or the arithmetically equivalent operation shown in Fig. 5). The **Proportion Image** is an 'attentional mask' that weighs 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 'A' that is potent to categorize 'A' as 'R'. Note that 'P' is the intersection of the represented information 'R' with the available information 'A'.

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% [4].

Bubbles for face identification

In high-level vision, we used *Bubbles* to understand the potent information underlying face identification [4]. To generate stimuli, we first decomposed the individual faces into six spatial frequency bandwidths of one octave each (see Fig. 6a and b). The bandwidth containing the coarsest information was used as a background. For 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. The 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).

'...*Bubbles* is a general technique that can assign the credit of a visual categorization task to its potent visual information.'

To derive $\mathcal{P}$ we computed a different Proportion Image per scale (Fig. 6c). For readability, we multiplied the scale information of Fig. 6b with the potent masks of Fig. 6c to obtain Fig. 6d. At the finest scale, the eyes and a corner of the mouth stand out (see the leftmost picture in Fig. 6d). 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 features could identify the face [5,39–41]. 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 Fig. 6e).

In summary, *Bubbles* is a general technique that can assign the credit of a visual categorization task to its potent visual information. In $\mathcal{R} \otimes \mathcal{A} \approx \mathcal{P}$, 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

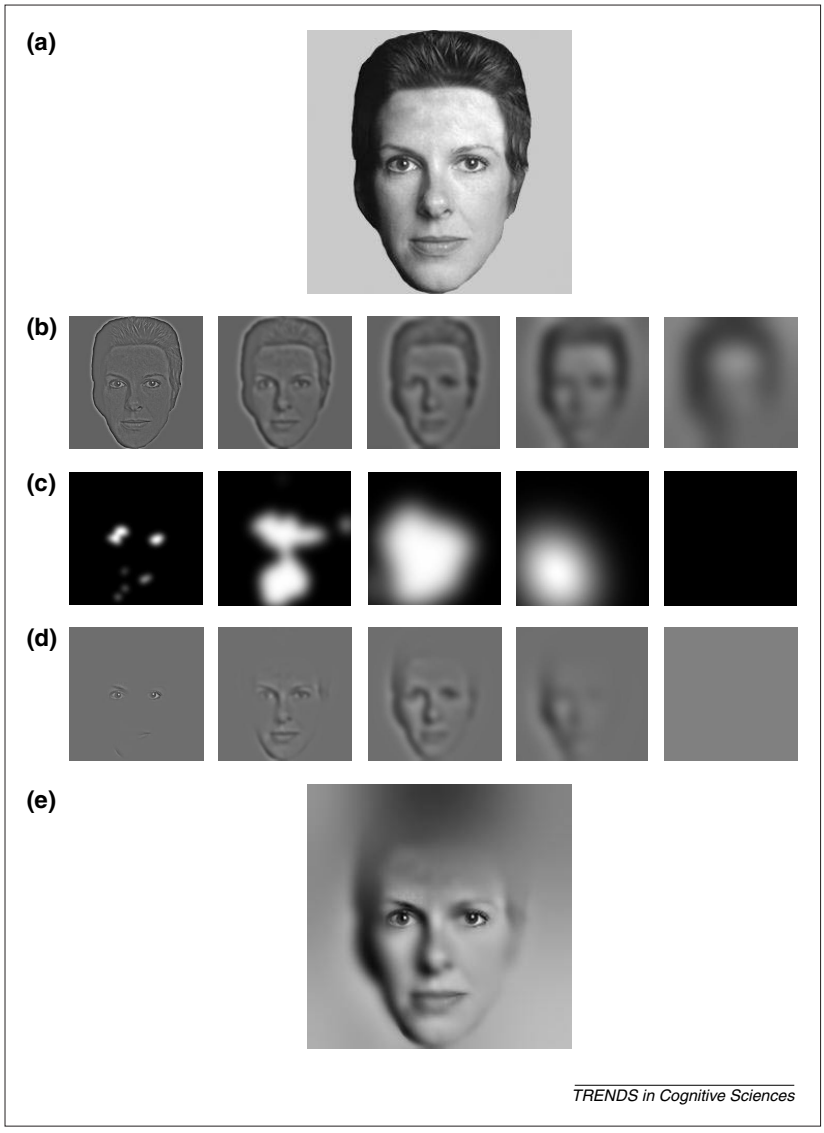

Fig. 6. The outcome of *Bubbles* in Experiment 2 of Ref. [4]. Participants learned 10 identities. They then attempted to identify the faces revealed by bubbles. The number of bubbles was adjusted to maintain the performance constant at 75% correct responses. Images in (b) represent five independent scales of (a) from fine to coarse: 90, 45, 22.5, 11.25, and 5.62 cycles per face. (c) Statistically significant potent regions at each spatial scale. (d) is the product of (b) multiplied by (c). (e) is the potent stimulus: a depiction of the information used to identify faces in Experiment 2.

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

$\mathcal{R} \otimes \mathcal{A} \approx \mathcal{P}$ can frame the information problem of visual categorization tasks. To be fruitful, $\mathcal{RAP}$ requires powerful handles on the nature of $\mathcal{R}$ (reverse correlation), $\mathcal{P}$ (*Bubbles*) and $\mathcal{A}$. What is $\mathcal{A}$, then? $\mathcal{A}$ is the information formally useful to resolve a specific task. To illustrate, imagine an observer in a toy world who must identify 10 images, which are equal in all points but for 100 varying grey-level pixels. Formally, only these 100 pixels form $\mathcal{A}$, because they are all useful to resolve the task. If the grey scale had 256 levels, A would be 100×8 bits of information.

If our observer had a perfect memory he could use each one of the 100 pixels to memorize the 10 images – for a total memory resource of $100 \times 8 \times 10$ bits of information. To this ideal observer, the information in memory would equal the information available to resolve the task, and $\mathcal{R} \otimes \mathcal{A} \approx \mathcal{P}$ would be $\mathcal{A} \otimes \mathcal{A} \approx \mathcal{A}$. Reverse correlation would reveal that $\mathcal{A} \otimes \mathcal{A} \approx \mathcal{A}$, and so would *Bubbles*, because each informative pixel can individuate the 10 faces for the ideal.

Human observers cannot represent the 10 images with all of the information in $\mathcal{A}$. Instead, they typically encode visual events with sparse representations that comprise only the most useful information in the task. A capacity-limited information bandwidth observer could isolate a subset of 10 pixels from the available 100, represent the 10 images in terms of these 10 pixels, and dismiss the remaining 90 available pixels. This information reduction would transform $\mathcal{A} \otimes \mathcal{A} \approx \mathcal{A}$ into $\mathcal{R} \otimes \mathcal{A} \approx \mathcal{P}$, to reflect that only 10 pixels of potent information represent the images in the memory of this observer (i.e. 10×8 bits of information per image).

From $\mathcal{A}$ and $\mathcal{P}$, we could measure the 'efficiency' [42] of the sparse observer in comparison to the ideal. In our example, the efficiency is related to $\mathcal{P}/\mathcal{A}$. Specifically, (10 images $\times$ 10 pixels $\times$ 8 bits)/(10 images $\times$ 100 pixels $\times$ 8 bits) = 0.1. Note that if the human observer was omniscient, its efficiency would become $\mathcal{P}/\mathcal{A} = \mathcal{A}/\mathcal{A} = 1$. Efficiency measures the proportion of available information that the observer can use to resolve a task.

To identify $\mathcal{R}$, we note that human observers have biases and beliefs that are independent of the information diagnostic in the task. Observers could believe that a certain number of pixels distinguish between the face when they do not. These pixels would neither be part of $\mathcal{A}$ (because they are constant across images), nor part of $\mathcal{P}$ (because useless pixels cannot be potent), but nevertheless be represented in memory, to represent contrasts with other categories. $\mathcal{R}$ then differs from $\mathcal{P}$ and $\mathcal{A}$, hence $\mathcal{R} \otimes \mathcal{A} \approx \mathcal{P}$.

Conclusions

This article has developed $\mathcal{RAP}$, 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 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 difficult dialogue between the domains of high- and low-level vision.

References

1 Backstein, K. (1992) *The Blind Men and the Elephant*, Scholastic
2 Bruner, J.S. and Goodman, C.C. (1947) Value and need as organizing factors in perception. *J. Abnorm. Soc. Psychol.* 42, 33–44
3 Helmholtz, H. (1856–1866) *Handbuch der Physiologischen Optik*, Voss
4 Gosselin, F. and Schyns, P.G. (2001) Bubbles: a technique to reveal the use of information in recognition tasks. *Vis. Res.* 41, 2261–2271
5 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
6 Wiener, N. (1958) *Nonlinear Problems in Random Theory*, John Wiley & Sons
7 Ahumada, A.J. and Lovell, J. (1971) Stimulus features in signal detection. *J. Acoust. Soc. Am.* 49, 1751–1756
8 Beard, B.L. and Ahumada, A.J. (1998) A technique to extract the relevant features for visual tasks. In *Human Vision and Electronic Imaging III, SPIE Proceedings* (Rogowitz, B.E. and Pappas, T.N., eds), Vol. 3299, pp. 79–85
9 Pfafflin, S.M and Mathews, M.V. (1966) Detection of auditory signals in reproducible noise. *J. Acoust. Soc. Am.* 39, 340–345
10 Ahumada, A.J. *et al.* (1975) Time and frequency analyses of auditory signal detection. *J. Acoust. Soc. Am.* 57, 385–390
11 Eggermont, J.J. *et al.* (1983) Reverse-correlation methods in auditory research. *Q. Rev. Biophys.* 16, 341–414
12 Sutter, E.E. and Tran, D. (1992) The field topography of ERG components in man: I. The photopic luminance response. *Vis. Res.* 32, 433–446
13 Simpson, W.A. *et al.* Identification of the eye–brain–hand system with point processes: a new approach to simple reaction time. *J. Exp. Psychol. Hum. Percept. Perform.* (in press)
14 Thomas, J.P. and Knoblanch, K. (1998) What do viewers look for when detecting a luminance pulse? *Invest. Opthalmol. Vis. Sci.* 39, S404
15 Barth, E. *et al.* (1999) Nonlinear features in vernier acuity. In *Human Vision and Electronic Imaging III, SPIE Proceedings* (Rogowitz B.E. and Pappas, T.N., eds), Vol. 3644, pp. 88–96
16 Neri, P. *et al.* (1999) Probing the human stereoscopic system with reverse correlation. *Nature* 401, 695–698
17 Watson, A.B. (1998) Multi-category classification: template models and classification images. *Invest. Opthalmol. Vis. Sci.* 39, S912
18 Watson, A.B. and Rosenholtz, R. (1997) A Rorschach test for visual classification strategies. *Invest. Opthalmol. Vis. Sci.* 38, S1
19 Gosselin, F. and Schyns, P.G. (2001) Superstitious perceptions. In *Proc. XIII Conf. Cogn. Sci. Soc.,* pp. 348–351, Lawrence Erlbaum
20 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
21 Jones, J.P. and Palmer, L.A. (1987) The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J. Neurophysiol.* 58, 1187–1211
22 DeAngelis, G.C. *et al.* (1993) Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex: I. General characteristics and postnatal development. *J. Neurophysiol.* 69, 1091–1117
23 Eckhorn, R. *et al.* (1993) The RF-cinematogram: a cross-correlation technique for mapping several visual receptive fields at once. *Biol. Cybern.* 69, 37–55
24 Emerson, R.C. *et al.* (1992) Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vis. Res.* 32, 203–218
25 McLean, J. *et al.* (1994) Contribution of linear mechanisms to the specification of local motion by simple cells in areas 17 and 18 of the cat. *Vis. Neurosci.* 11, 271–294
26 Oshawa, I. *et al.* (1990) Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science* 249, 1037–1041
27 Ringach, D.L. *et al.* (1997) A subspace reverse correlation technique for the study of visual neurons. *Vis. Res.* 37, 2455–2464
28 DeAngelis, G.C. *et al.* (1995). Receptive–field dynamics in the central visual pathways. *Trends Neurosci.* 18, 451–458
29 Gold, J. *et al.* (2000) Deriving behavioral receptive fields for visually completed contours. *Curr. Biol.* 10, 663–666
30 Mangini, M.C. and Biederman, I. (2001) Differentiating expression, gender, and identity in faces: comparing normals, the ideal observer, and a prosopagnosic. *Proc. Vis. Sci. Soc.* 1, 90
31 Abbey, C.K. *et al.* (1999) Estimation of human-observer templates in two-alternative forced-choice experiments. In *Proceedings of the Society of Photo-optical Instrumentation Engineers*, (Krupinski, E.A., ed.), pp. 284–295, SPIE
32 Eckstein, M.P. *et al.* (2001) The footsteps of attention in the Posner paradigm revealed by classification images. *Proc. Vis. Sci. Soc.* 1, 26–27
33 Hubel, D.H. and Wiesel, T.N. (1959) Receptive fields of single neurons in the cat's striate cortex. *J. Physiol.* 148, 574–591
34 Hubel, D.H. and Wiesel, T.N. (1962) Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154
35 Ringach, D.L. and Shapley, R. (1996) Spatial and temporal properties of illusory contours and amodal boundary completion. *Vis. Res.* 36, 3037–3050
36 Bayer, H.M. *et al.* (1998) Recognizing facial expressions efficiently. *Invest. Ophthalmol. Vis. Sci.* 39, S172
37 Solomon, J.A. and Pelli, D.G. (1994) The visual filter mediating letter identification. *Nature* 369, 395–397
38 Pelli, D.G. *et al.* Identifying letters. *Vis. Res.* (in press)
39 Gauthier, I. and Tarr, M.J. (1997) Becoming a 'Greeble' expert: exploring the face recognition mechanism. *Vis. Res.* 37, 1673–1682
40 Sergent, J. (1986) Microgenesis of face perception. In *Aspects of Face Processing* (Ellis, H.D. *et al.,* eds), Martinus Nijhoff
41 Tanaka, J. and Sengco, J.A. (1997) Features and their configuration in face recognition. *Mem. Cognit.* 25, 583–592
42 Kersten, D. (1990) Statistical limits to image understanding. In *Vision: Coding and Efficiency* (Blakemore, C., ed.), pp. 32–44, Cambridge University Press