

# Eidolons: Novel stimuli for vision research

**Jan Koenderink**

Justus-Liebig Universität Giessen,  
Abteilung Allgemeine Psychologie, Giessen, Germany  
KU Leuven, Laboratory of Experimental Psychology,  
Leuven, Belgium  
Utrecht University, Department of Experimental  
Psychology, Utrecht, The Netherlands

**Matteo Valsecchi**

Justus-Liebig Universität Giessen,  
Abteilung Allgemeine Psychologie, Giessen, Germany

**Andrea van Doorn**

Justus-Liebig Universität Giessen,  
Abteilung Allgemeine Psychologie, Giessen, Germany  
Utrecht University, Department of Experimental  
Psychology, Utrecht, The Netherlands

**Johan Wagemans**

KU Leuven, Laboratory of Experimental Psychology,  
Leuven, Belgium

**Karl Gegenfurtner**

Justus-Liebig Universität Giessen,  
Abteilung Allgemeine Psychologie, Giessen, Germany

Meanings and qualities are fundamental attributes of visual awareness. We propose “eidolons” as a tool for establishing equivalence classes of appearance along meaningful dimensions. The “eidolon factory” is an algorithm that generates stimuli in such a meaningful and transparent way. The algorithm allows us to focus on location, scale, and size of perceptually salient structures, proto-objects, and perhaps even semantics rather than global overall parameters, such as contrast and spatial frequency. The eidolon factory is based on models of the psychogenesis of visual awareness. It affects the image in terms of the disruption of image structure across space and spatial scales. This is a very general method with many potential applications. We illustrate a few instances. We present results for the example of tarachopic amblyopia, showing that scrambled vision is indeed an apt interpretation.

## Introduction

Wouldn't it be nice to experience what colorblind people are seeing or what tarachopic amblyopes have to cope with? Viénot, Brettel, Ott, M'Barek, and

Mollon (1995) told us about the former by simulating the visual appearance of unilateral dichromats. Hess (1982) informed us about the latter by requiring patients to draw what they see. This is by no means trivial. Unilateral dichromats are not all that clear in their reports (Sloan & Wollach, 1948), whereas they surely should be the first to know! It becomes clear that such introspective reports are less easy to come up with than it might seem if you try to describe—even to yourself—what you experience as you see a book shelf in the periphery of your visual field. You may somehow be aware of the presence of books with titles written on their spines, yet you cannot identify the books or read the titles. The perceptual quality of peripheral vision appears contradictory and defies your ability to describe your sensations, yet surely *you* are the first to know. Perhaps Titchener's (1902) lab manual should be consulted; he explains in detail how to use introspection.

Phenomena such as visual crowding have been studied by measuring detection and discrimination thresholds. The appearance of suprathreshold stimuli is much harder to study. Exceptions are very specific situations. One instance is direct contamination be-

Citation: Koenderink, J., Valsecchi, M., van Doorn, A., Wagemans, J., & Gegenfurtner, K. (2017). Eidolons: Novel stimuli for vision research. *Journal of Vision*, 17(2):7, 1–36, doi:10.1167/17.2.7.

doi: 10.1167/17.2.7

Received January 20, 2016; published February 28, 2017

ISSN 1534-7362

This work is licensed under a Creative Commons Attribution 4.0 International License.



Downloaded From: <http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/936040/> on 12/13/2017

tween flanker and target objects (Greenwood, Bex, & Dakin, 2010). One really needs first-person reports—for instance, having observers draw what they see (Metzger, 1936; Sayim & Wagemans, 2013). However, the (in)ability of observers to reproduce their visual experience limits this technique to relatively simple stimuli. Another way is to use verbal descriptions (for review, see Lettvin, 1976; Metzger, 1936; Pelli, 2008). Of course, the use of first-person reports is beset with difficulties.

Another approach might take its lead from methods as proposed by Viénot et al. (1995), who transformed color images to emulate dichromatic vision. Why not preprocess a stimulus in ways that emulate peripheral processing and present it foveally? Indeed, such methods have been proposed and used to good advantage by authors such as Rosenholtz and colleagues (e.g., Balas, Nakano, & Rosenholtz, 2009). In such cases, one really needs to test various methods of preprocessing to mimic the peripheral data. This might (at least) serve to generate various hypotheses about what goes on in peripheral vision as viable or perhaps worth pursuing.

Such methods have considerable potential. However, in order to wield them effectively, one needs to be able to explore a reasonable environment of the original stimulus. Because images can be transformed into other images in infinite ways, whereas empirical research can explore only limited ranges, there is a need for controlled variation based on our present understanding of visual processes. This is by no means an understood issue and has remained more of an art (Balas & Conlin, 2015). One really needs a much more transparent and intuitive way of parameterizing and perturbing stimuli.

Here we introduce a novel processing algorithm that produces stimuli differing from a given image in ways that are controlled by a limited number of clearly understandable parameters. Within this parametric space, we refer to the subset of stimuli that are equivalent along a given perceptual domain as *eidolons* (see Appendix D).<sup>1,2</sup>

Of course, one could use a very simple parameter space. The variable contrast of sine-wave gratings used in a modulation transfer (MTF; Schade, 1948, 1956) measurement draws on a one-parameter space of eidolons—the parameter being Michelson contrast. On the other side one could use a very complex parameter space to the point of parameterizing the luminance of every pixel within an image. At the same time, one could use a very strict criterion to establish perceptual equivalence (e.g., defining two stimuli as equivalent only when they are metameric) that is by all means indistinguishable or a wider sense equivalence criterion based, for instance, on semantics. In this wider sense, all the well-known—even famous—instances of Leo-

nardo's Mona Lisa by Marcel Duchamp, Fernando Botero, and many other artists are eidolons.

In experimental phenomenology, one aims at parameterizations that naturally fit generic visual presentations rather than imposed physical ones. An example of two physical parameters that do not map to perception is contrast and sharpness in images. As photographers know, a high-contrast print often serves to save a slightly unsharp shot. Likewise, low-contrast prints are often considered unsharp. In such cases a natural space of eidolons may be a useful platform from which to launch research.

Another potential use of eidolons is in the study of visual anomalies and agnosias. Well-known examples are renderings of color photographs that are intended to suggest to the normal trichromat what the experiences of various dichromats might be like. While this use of eidolons might not directly answer scientific questions, such examples are useful because they offer the generic observer an opportunity to better understand more or less singular ones and thus interact more effectively with them. For instance, one might adapt one's printed or projected figures and text for more universally effective communication. Thus, the topic holds some genuine interest, even though mostly from an applied science point of view.

The body of this article is structured in three sections. Theory details the theoretical framework in which we define our concept of eidolon. Implementation contains the general description of an eidolon factory based on scale decomposition and spatial disarray. Examples shows some examples of experiments in which stimuli produced by the eidolon factory could be used.

## Theory

### Formal description of eidolons

As anticipated, we define an eidolon of an image as the equivalence class of images that evoke the same visual awareness—given certain specified constraints—in a given observer as the fiducial image.<sup>3</sup> This implies that the eidolon may formally represent that awareness or aspect of awareness. An equivalence relation in vision is commonly based on whether two images look alike. In empirical science, this definition is unmanageable and needs to be converted to some operational form. This involves two distinct aspects. One is the operationalization of the equivalence. This boils down to one or more formal psychophysical measures. The other is the description of the equivalence set. This may take the form of an algorithm that produces instances on call or a set of parameters that allows attaching a



Figure 1. Two Julesz patterns. Each has  $100 \times 100$  square patches randomly colored white or black with 50% probability. The set of  $2 \times 10^{3010}$  images contains such instances as a rendering of the Mona Lisa, Jackson Pollock drawings, and so forth, yet the overwhelming number of them all look the same. The set is much more effectively described by a simple algorithm than by listing all members. To most observers, at a brief glance, both instances will look identical. There may be the occasional eidetic (sometimes described as a savant skill among individuals with autism spectrum disorder), but we know little about their actual abilities.

unique formal label to an instance. The algorithm might be deterministic or stochastic. Likewise, the parameterization might be deterministic or stochastic.

A well-known example of an algorithm to produce instances of equivalent images is the familiar JPEG file format. Remember that JPEG (Pennebaker & Mitchell, 1993) uses lossy compression—that is, it literally replaces the fiducial image with a phantom lookalike. The user never notices losing the original except in cases of extreme compression. The JPEG eidolons have become of great economic importance and are in common use.

Perhaps the best-known case of equivalence classes in human vision is color. Infinitely many spectral compositions look like the same baby blue. Even close scrutiny in metameric lights does not reveal differences. A color is in fact a huge class of physically distinct radiations that are phenomenologically identical (Koenderink, 2010). All metameric stimuli are trivially eidolons in our definition. Conventional display units present one with such chromatic eidolons.<sup>4</sup> The space of spectral compositions they offer has measure zero in the space of physical spectra yet is sufficiently extended given the bottleneck of human physiology.

An eidolon may have very high cardinality. Consider the early Julesz patterns (Julesz, 1962; see Figure 1)— $100 \times 100$  pixel 1-bit images in which each pixel was white or black with probability one half. Define equivalence as the inability to keep two images apart at cursory view. Then any such image, with probability one, is part of the eidolon, for no two instances look different.<sup>5</sup> That is why you generically cannot describe an eidolon by exhaustively listing its members. De-

scription by algorithm—the algorithm being controlled by a few parameters—is the only operationally viable solution.

Other parametric families of images can be generated by manipulating physical aspects of the scene being depicted, including geometrical transformations such as translation and rotation of objects, size manipulations (e.g., Biederman & Gerhardstein, 1993; Logothetis & Sheinberg, 1996), and/or changes in the illumination in terms of direction, intensity, and spectral composition (e.g., Hurlbert, 2007; Kraft & Brainard, 1999). These are of obvious ecological relevance, and fruitful lines of research have shown that the visual system builds neural representations that disentangle the information relative to the objects being represented and to the transformations themselves (e.g., DiCarlo & Cox, 2007; Hong, Yamins, Majaj, & DiCarlo, 2016; Rust & DiCarlo, 2012). Families of stimuli can also be generated in order to contrast different models of the visual system (see Wang & Simoncelli, 2008) and/or their parameter values. One of many examples is the manipulated scenes used by J. Freeman and Simoncelli (2011), which were predicted to be perceptually equivalent under the assumption of a given scaling of receptive field size as a function of eccentricity.

Many eidolon factories are thus possible. The eidolon factory we propose is a class of methods that allows easy generation on demand using simple, intuitive parameterization. In this article we also describe how to set up one's own eidolon factory for vision research. In Examples we illustrate the use of such methods to explore the topic of tarachopic amblyopia (Hess, 1982). Of course, this is no more than a proof of principle. We also provide a few examples of possible experiments that might be conducted using eidolons, although we can only scratch the surface of possibilities.

We don't claim to explain or model the brain with these eidolons; we simply forge a phenomenological model that we believe will prove useful in experimental phenomenology (Albertazzi, 2013; Koenderink, 2015a). However, we do consider the deep structure of the visual field—that is, the structure that recognizes both scale and spatial variation in the design. We use methods of a geometrical nature that will be recognized as having considerable similarity to structural descriptions commonly encountered in the formal theories of neural architecture. We develop this in the following sections.

## A model of psychogenesis

Objective psychophysics is concerned with threshold measurements, ignoring visual qualities and meanings. Modeling is primarily based on aspects of physics



(mainly optics) and physiology. In the case of the eidolon factory, one has to use a model of psychogenesis instead because we are dealing with visual awareness in the experimental phenomenology of the sense of sight.

Our model is based on the notion that awareness is a mental construction. This idea was prominently put forward by Hermann von Helmholtz (1892) in his suggestion that perception is largely the result of unconscious inferences about the world, and it is now commonplace that perception is at least in part constructive (Gregory, 1997; Hochstein & Ahissar, 2002; James, 1890).

In models of the brain the visual cortex is often constructed as representing optical data, or even scene data. In our model we regard the neural representation as similar to the representation offered by the moist sand of a beach. A depression might be said to represent the impression of a bare human foot. But notice that the sand—being inanimate—must be fully oblivious of its representing anything. Likewise, the cortical representation is meaningless as far as awareness is involved (Koenderink, 2015b). It is very useful as a structure that constructs a volatile (because continually overwritten) file that records the available optical structure. Like the file of a forensic investigation, most of what is in the file will never even be consulted. Awareness is constructed and perhaps originates from dreamlike states. In psychogenesis, such states diversify and evolve, eventually attempting to account for the activity in the primary visual areas. This is how the hallucinations are constrained. Thus, the cortex serves in sculpting imagery so to speak.

A full discussion of such a model would go too far here. We refer to our recent article for a more detailed description (Koenderink, van Doorn, & Pinna, 2015). We merely give one example: the case of edges (Marr & Hildreth, 1980; Savant, 2014). Suppose an input image comprises two uniform areas of different intensities meeting at a straight common boundary. Such a configuration is considered an ideal edge. First-order directional derivatives of any scale will represent the edge as a straight, fuzzy (Zadeh, 1965) ribbon of activity, its width being determined by the scale. Psychogenesis will construct the appearance of an edge by painting the visual field with a row of edgelets to account for the representation. The edge would appear light on one side and dark on the other side—a bit like Pinna's watercolor illusion (Pinna, 1987, 2008; Pinna, Brelstaff, & Spillmann, 2001; Pinna, Werner, & Spillmann, 2003). This is very unlike the edginess represented in the neural structures. A formal analysis shows that the edgelet presentation in visual awareness at a given scale equals the Laplacian of the input pattern (Koenderink et al., 2015; Koenderink, van Doorn, Pinna, & Wagemans, 2016). Summing over all scales

yields the input pattern except for its average level. Thus, we recognize two distinct levels: the meaningless and qualityless neural representation (the analysis) and the construction of visual awareness, which is a creative imagery (the synthesis).

In the eidolon factory, we implement exactly such analysis and synthesis stages. The variability in the eidolon instances is due to the synthesis, the analysis being a straightforward algorithm. This mimics psychogenesis, which works well with lacunary data (Kanizsa, 1997; Metzger, 1936), much like scientific observation in many fields (Monmonier, 1999). Visual awareness also easily deals with and even prefers local disarray (Gombrich, 1963; Weegee & Speck, 1964) or fuzziness (see Vasari, 2007).

## The structure of the visual field

Evidently, eidolon algorithms should respect the structure of the visual field, which is a mental entity.<sup>6</sup> If one desires to arrive at simple, intuitive designs, it makes sense to rely more on basic, formal geometrical principles than on our current understanding of the neurophysiology of the primary visual cortex (Hubel & Wiesel, 1968; Kandel, Schwartz, & Jessell, 2000). That is where formal descriptions have to come from. Moreover, we look primarily at the phenomenology of visual awareness rather than neurophysiology.

A first principle might be that a not-too-large patch of the visual field is approximately uniform and isotropic in its properties. Here we ignore the global—though major—effect of eccentricity; the remark applies to local structure. The visual field has a structure focused on rather local regions. It makes solid ecological sense. Generically, what happens here is quite distinct from what happens there. Local structure tends to be statistically uniform.

This focus on locality rules out methods such as global Fourier analysis (Papoulis, 1962), although this method is widely used in vision research. This holds equally for other global descriptions. What renders Fourier analysis special is its translation invariance, which is indeed a desirable property.

Fourier analysis is linear. This is another desirable property. You may decompose any image in Fourier components and synthesize it back again. Perhaps unfortunately, this has led people, especially in vision research, to infer that images *comprise* Fourier components (Maffei & Fiorentini, 1973). This inference is mistaken because linearity fully hides the nature of composition after synthesis. You will never know whether “4” was synthesized as “2 + 2” or “1 + 1 + 1 + 1.” Linear vector spaces of any dimension lack natural parts.

The natural parts issue is of fundamental importance. Is a sausage composed of slices? Of course not! Yet you can certainly *decompose* a sausage into slices. But that such a decomposition is possible by no means implies that the whole is *composed* of the (resulting) parts. A sausage can be cut in many ways; thus, sausage slices are by no means natural parts. Fourier methods—the basis of the MTF method—imply parts (so-called sine-wave gratings) that are just as arbitrary as the sausage slices.

Fourier methods are *global*. At the opposite side of the range, one has purely *local* methods. Global versus local is another fundamental distinction. The formal method to describe local geometry is differential geometry (Spivak, 1999). Here one encounters considerable affinity with many familiar properties of the visual field (Koenderink, 1984b, 1990; Koenderink & van Doorn, 1992). The visual field is structured as a hierarchy of localities.

Differential geometry exploits the structure of the Euclidean plane in the immediate (so-called infinitesimal) neighborhood of any of its points (Bell, 2005). Because all such neighborhoods are considered mutually congruent, one has again a translationally invariant, linear structure. However, this requires a connection, or glue mechanism, that allows one to compare points that are sufficiently close by. A connection is a formal mechanism that allows transport of geometrical structure from one location to another (Levi-Civita & Ricci, 1900). The simplest geometrical structures that need to be transported in order to become fit to mutually compare are tangent vectors. A vector is the derivative of a point, or, equivalently, a bilocal object. It has a direction and a magnitude (Koenderink, 1990; Koenderink & van Doorn, 1992; Koenderink et al., 2015). All vectors at a point span the tangent plane, which is a representation of the infinitesimal neighborhood of the point. The connection relates close-by tangent planes to each other.

Notice how this has an obvious likeness to the neurophysiological concept of a small set of receptive fields overlapping at some point (Koenderink, 1984a). All that is lacking is the notion of the size of a point. According to one of Euclid's definitions, a point is "that which has no parts." Euclid (ca. 300 BCE; see Burton, 1945) nowhere says points need to be small. Indeed, if one considers size invariance of visual field properties important, one should consider points of any size (Kandinsky, 1959). Thus, the visual field should have a self-similar structure (Koenderink, 1984b), lacking an absolute measure of size.<sup>7</sup> Technically, this implies linear scale space, now a standard tool in image processing (Burt & Adelson, 1983; Florack, 1997; Koenderink, 1984b; Lindeberg, 1994; Schmalzing, 1997; ter Haar Romeny, 2008). The great advantage is its simplicity; it is just differential geometry augmented with size.

Many operators in common use are much more complex than their scale-space equivalents. But that comes at a price: a loss of generality. For instance, there are many edge detectors that find useful applications in specific tasks. But edges are in the mind, not in the image, and they come in great variety (Koenderink et al., 2015), as painters well know (Cateura, 1995; Jacobs, 1986). No specific edge detector is optimal in all circumstances because they have necessarily been optimized under certain prior assumptions.<sup>8</sup>

In the final analysis, only differential geometry makes general sense exactly because it is about nothing in particular. We have explained the case of edges in some detail elsewhere (Koenderink et al., 2015, in press). In the differential geometric formalism, edges are an alternative for points in that they provide a complete, linear representation of the image. This is categorically different from the mainstream view, where edges are singular features that allow partial reconstruction of an image due to smart sparse coding (Elder & Zucker, 1998; Marr & Hildreth, 1980).

Something similar holds for receptive field profiles. Would the eidolon factory be better off with Gabors (Gabor, 1946)? Well, that would certainly complicate the mathematics greatly without any actual gain.<sup>9</sup> Maybe it makes one feel better (closer to the physiology, maybe) that Gabors have numerous zero crossings all the way to infinity? Who knows. From a formal and algorithmic perspective, they are an unnecessary headache. There are some indications that the scale-space profiles are okay models of the physiology too (Lindeberg, 2013; Young, 1987). If so desired, an eidolon factory could be designed on a basis of Gabor analysis/synthesis too.

Scale space is the model of the visual field we will use (Koenderink, 1984b; ter Haar Romeny, 2008). It is local, isotropic, translation invariant, and self-similar. Its local structure is differential geometry with structure up to some low order—at least two. This induces a receptive field structure that is very reminiscent of what is described for the primary visual cortex (Koenderink, 1990; Koenderink & van Doorn, 1992; ter Haar Romeny, 2008). It is also a formal system that allows geometrical operations in a transparent and exact manner.

## Local sign

The notion of "local sign" (meaning "positional signature"; *Localzeichen* in German<sup>10</sup>) is due to Hermann Lotze (1852). Although it was considered of major importance at the time (mid-19th century) and frequently occupied people such as Helmholtz (1892), it has largely been forgotten today. In order to appreciate what the local sign problem is, consider the following.

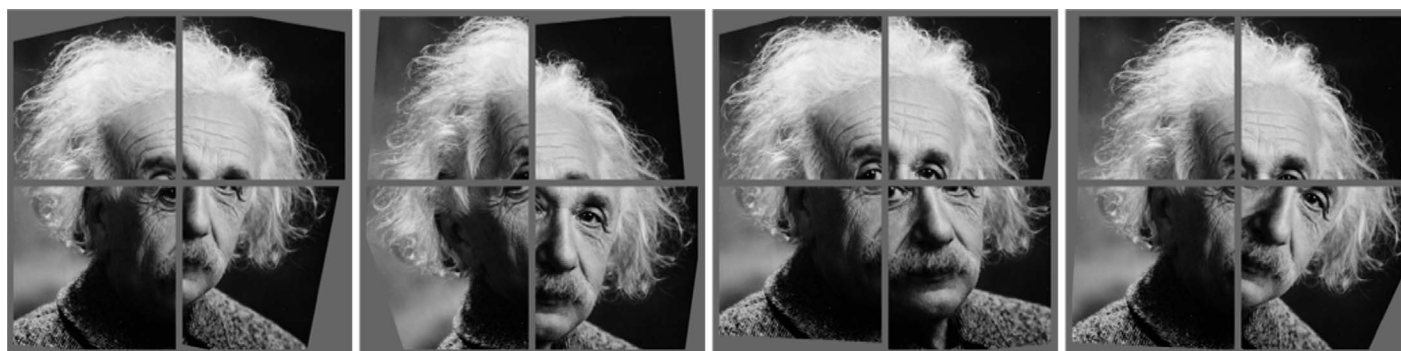


Figure 2. Here are some fairly extreme independent translation–rotation offsets relative to each other of the quadrants of a square image. When viewed briefly, these are just faces; they are somehow merged. A closer scrutiny can reveal the spatial disarray. In peripheral vision, they are perceptually equivalent and are part of a single eidolon. Even in focal scrutiny, quite large disarrays are easily missed. Even when they are noticed—as in these examples—one somehow experiences a fairly coherent impression. The illusion of image coherence becomes even stronger at larger separation. Photograph of Albert Einstein reprinted from <http://www.loc.gov/pictures/item/2004671908/>. Copyright Orren Jack Turner.

Axons carry spike trains, one action potential looking much like the other. A recording from any given axon will fail to reveal which location of the visual field the neuron is serving, what the specific property the neuron is messaging about might be, what its preferred direction or orientation is, what the current gain factor of the neuron is, and so forth. Of course, the brain scientist knows. This is possible because the scientist knows the current stimulus, has poked an electrode at a certain location, can look at the anatomy, and might have additional information concerning the neuron. But how about the brain itself? It does not know the stimulus, it cannot look at its own anatomy, and it is unlikely to maintain databases on all of its neurons.

Modern brain scientists consider the discovery of somatotopy to have rendered Lotze’s problem a nonissue<sup>11</sup> (Kaas, 1990). We hold a rather different opinion on this and believe that the problem might be even more pressing than it was felt to be at the time.

There are indications that an apparently physiologically intact visual brain might still lack local sign. We refer to the condition of *tarachopia* (“scrambled vision”) as reported by Hess (1982). In a case of unilateral tarachopia, the tarachopic eye might have acuity and contrast sensitivity just as good as that of the normal eye, yet the amblyopia might be serious enough to prevent the owner making out today’s newspaper’s headlines. Apparently the tarachopia involves lacking or disturbed local sign. The local structures are there, but the differential geometric connections are lacking. The physiological basis is (as yet) unknown, so to this day tarachopia had to be classified as an agnosia (*Seelenblindheit*, or “soul blindness”). This shows that local sign is an ill-understood mechanism that lies at the basis of visual awareness and is not a mere philosophical fiction, as is sometimes suggested.

We will not enter into the possible mechanisms of local sign here (see elsewhere—e.g., Koenderink, 1984a), but we suggest that scrambling local sign, or local disarray of image elements, might be an apt model of certain forms of visual equivalences. For instance, the phenomenon of crowding in the peripheral visual field (Bouma, 1970; Pelli, 2008) is phenomenologically very similar to the tarachopic condition as it occurs in the focal vision of certain patients (e.g., Hess, 1982; Sayim & Wagemans, 2013).

From a phenomenological perspective, at least some aspects of local sign appear to be implemented on the fly in the psychogenesis of visual awareness. If one cuts an image into pieces and displaces the pieces randomly, the disarrayed image looks fairly normal (Figure 2). Masking the seams between the pieces (e.g., with gray stripes) results in the experience of an undisturbed image, seen behind the mask (Koenderink, Richards, & van Doorn, 2012a). This also works in space time (Koenderink, Richards, & van Doorn, 2012b). In such cases the optical data are incoherent, whereas the visual awareness is coherent—a most remarkable fact! Apparently, the perception constructs an orderly image where physically there is chaos. These effects work over large distances and time spans. They remain ill (euphemism for “not at all”) understood.

Such empirical facts suggest that the psychogenesis of visual awareness imposes spatiotemporal coherence in an active way. Possibly classical local sign might—at least partly—depend on this. The crowding phenomenon (Bouma, 1970; Pelli, 2008) further suggests that such a mechanism is confined to focal vision. We expect the eidolon factory to become an important tool in the study of such phenomena.





Figure 3. Some samples from a scale space. The scales are 1.2, 9.5, 20, and 59 compared with  $512 \times 512$  for the full image. Notice that blurring simplifies the image. The leftmost image might be taken for the fiducial; the difference would be invisible in print. Photograph of angel statue reprinted from <https://pixabay.com/en/bust-sculpture-statue-fig-art-1555688/>.

## Implementation

### Methods

We use pseudocode to present our descriptions in an algorithmic language (Appendix B). We also make available a demolike implementation written in Processing (Reas & Fry, 2010; Appendix A) that should run without much ado on most common platforms. A Matlab implementation is available at <http://www.allpsych.uni-giessen.de/EidolonFactories/index.htm>. A Python implementation is available at <https://github.com/gestaltrevision/Eidolon>.

### The basic deterministic structures

Every factory run necessarily starts with a fiducial image. This image is the basis for a huge data structure that is perhaps best thought of as a simulation of the cortical activity induced by that image. This is the structure that will eventually be used in the synthesis. Once set up, the fiducial image itself has become irrelevant. This is not so much an analysis as merely a dumb formatting stage.

Consider an image as a discrete sample of a scalar field (intensity say) defined over the Euclidean plane. The first chore is to represent it at multiple levels of resolution. This is the proper topic of scale space (Florack, 1997; Koenderink, 1984b; Lindeberg, 1994; ter Haar Romeny, 2008). It has long since become a de facto standard in image processing. In practice, one samples both the scale and the space domain discretely.

In order to understand the scale-space data structure one needs a number of important insights. These are all based on the basic scale-space structure. Although we will not prove it here (textbooks quoted previously), scale-space is based on the Gaussian kernel as point. A convenient scale parameter is the half-width (standard

deviation) of the Gaussian blurring kernel. In Figure 3 we show the (sampled) scale space for an image.

Because the image at some scale can be computed from any image at some finer scale, this representation is inconveniently—and quite unbrainlike—redundant. It is preferable to find the *differences* between adjacent scale levels. A stack of such difference images will be our basic data structure. It represents the simplest possible nontrivial structure that has the desirable properties listed above: local, isotropic, translation invariant, and self-similar.

The difference layers are just the fiducial image as represented in difference of Gaussian (DOG) receptive fields of various sizes (Figure 4). Another way to understand these layers is as a stack of scale derivatives of the image. This explains the use of DOG filters in sharpening images (Margulis, 1998, 2005) in applications such as Adobe Photoshop (San Jose, CA).

The implication is that all layers added together simply recompose the image. This is indeed obvious from the differences definition. But although trivial, this is a crucial point. The eidolon factory analyzes the image and synthesizes it again. If done exactly like this, you would construct the perfect doppelgänger—namely, the picture itself! The desirable fuzziness of the eidolons derives from perturbations applied to the parts before the synthesis. This is the eidolon factory in a nutshell.

Yet another interpretation is important. The differences show local transition regions. This is different from edge finder images in that the local transitions have sides, whereas mere edginess—as yielded by programs such as Photoshop—fails to represent that. This is an important topic. It suggests that the integration over DOG activity can also be understood as a synthesis that combines all edgelet (or transition area) samples. This can indeed be proven formally (Koenderink et al., 2016). It is somewhat intricate because the argument involves edge finders, line finders (directionally and orientationally tuned “simple cells”; Figure 5; Hubel & Wiesel, 1968), and Laplacian

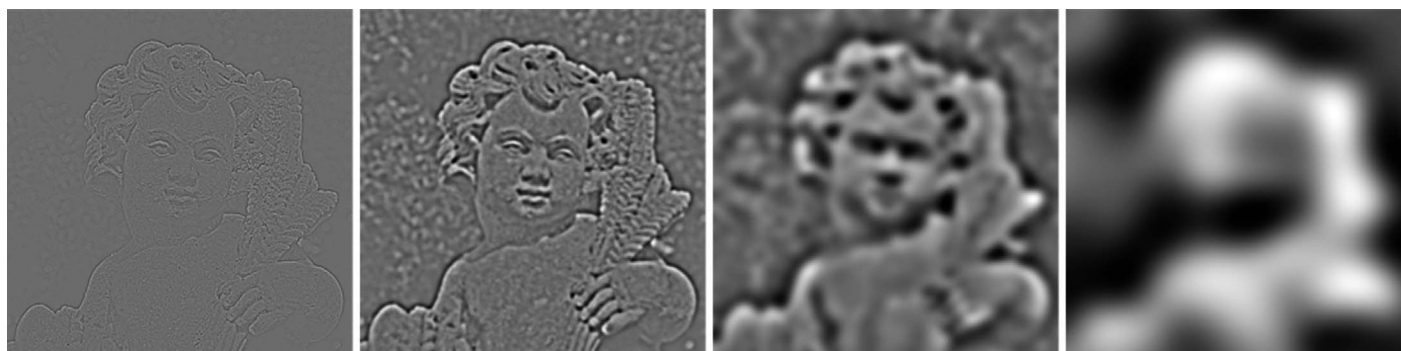


Figure 4. Some layers from the DOG scale space. Each layer carries structure of a given scale; adding all layers together reproduces the fiducial image. Photograph of angel statue reprinted from <https://pixabay.com/en/bust-sculpture-statue-fig-art-1555688/>.

operators. It allows one to use the DOG responses as summaries of the (far more numerous and involved!) simple cell responses. For details we refer to earlier work (Koenderink, 1990).

## Local disarray

The notion of local disarray is vital to the eidolon factory because it yields the slop or fuzziness that renders the eidolons extensive.<sup>12</sup> Integrating the scale-space difference layers (or DOG activities) simply reproduces the fiducial image. In that sense the eidolon factory is like global Fourier analysis: It is complete and linear. The crucial difference is that it is *local*, so we can introduce local perturbations. Notice that there are infinitely many local bases possible. We stick to the simplest, which is just differential geometry with scale.

A perturbation of local sign can conceptually work out in many different ways. Consider two extreme cases. In the first case, if you add an identical offset to

the local sign of every receptive field, you simply shift the image. Such a fully coherent perturbation has no visual effect. In the second case, if you add a statistically independent offset to the local sign of every receptive field, you destroy the image. Such a fully incoherent perturbation has a visual effect that is similar to—but by no means the same as—blurring.

It is important to understand that whereas disarray and blurring both change the effective resolution, they are nevertheless essentially different. The difference is that disarray preserves the (perhaps local) histogram. Thus, if you thoroughly blur a chessboard image you obtain an average uniform gray image, but if you thoroughly disarray a chessboard image you obtain a random pixel array in which pixels are white or black with probability one half. Whereas the blurred version is uniformly gray, the disarrayed one is simultaneously white and black. They are by no means visually equivalent (Figure 6). In the disarrayed state, “white and black” is a bona fide color.

Depending on the precise style of disarray, one obtains very different results (Figure 7). One might expect a parameterization to be really complicated.

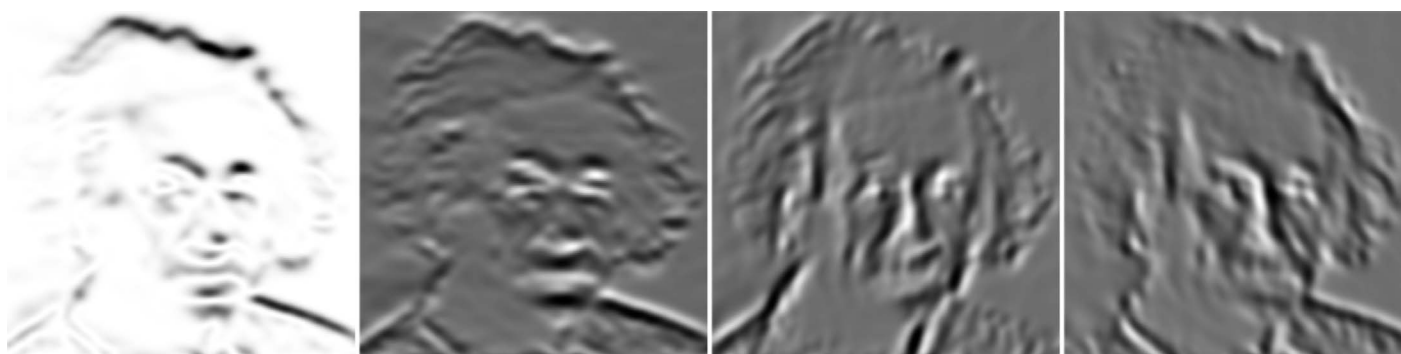


Figure 5. At left edginess, that is r.m.s. first-order (edge finder) activity over directions. The three other images show second-order (line finder) activity for orientations of 90°, 150°, and 30°. These three represent all orientations exactly. Adding all line finder activity at some scale yields the DOG activity at that scale; adding all DOG activity reproduces the image. Notice how the barcode for vertical variation (second image from the left) neatly represents the generic structure of a human face. It is a scheme used by draughtsmen throughout the centuries (Koenderink et al., 2016). Photograph of Albert Einstein reprinted from <http://www.loc.gov/pictures/item/2004671908/>. Copyright Orren Jack Turner.



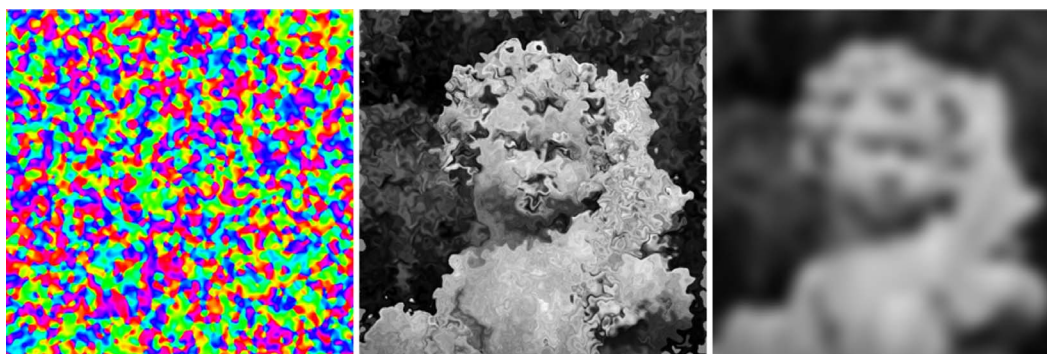


Figure 6. Left: a Gaussian random two-dimensional vector field (hue indicates direction); center: an image that was locally disarrayed by this field; right: an image blurred to about the same effective resolution. The effects of disarray and blur are very different. Disarray conserves the histogram and blurring does not, so the blurred image has lost contrast. Fine details of the disarrayed image are spurious, whereas the blurred image remains fully veridical, only less detailed. Photograph of angel statue reprinted from <https://pixabay.com/en/bust-sculpture-statue-fig-art-1555688/>.

This is an important topic when eidolons are to be used in vision research. A simple, intuitive parameterization is a *sine qua non* for focused investigation to be possible at all. An eidolon factory that is intuitively opaque and depends on complicated parameterization would be like a sledgehammer, whereas the desirable tools are tweezers and a scalpel. We consider the important issue of parameterization in the next section.

### Eidolon parameterization

In perturbing local sign, we identify three conceptually different handles. These relate to the structure of the disarray fields. We denote these *reach*, *grain*, and *coherence*. We discuss each in sequence. All are based on disarray generated by Gaussian random fields that are statistically uniform and isotropic. Of course, it is

possible—and sometimes desirable—to consider non-uniform and/or nonisotropic perturbations. This is perfectly possible (we show an example later), but one needs then to forge a suitable parameterization that applies to such special cases.

Basic Gaussian random fields are easily obtained by blurring Gaussian white noise. Generating two independent instances at the same blur level and with the same power yields Gaussian random displacement vector fields. One simply treats the scalar fields as the Cartesian coordinates of displacement vectors. Here we use the fact that an isotropic normal distribution in the plane is separable into two mutually independent scalar normal distributions. This is an important, highly remarkable property that renders the normal distribution unique. In Figure 8 we illustrate such random displacement fields.

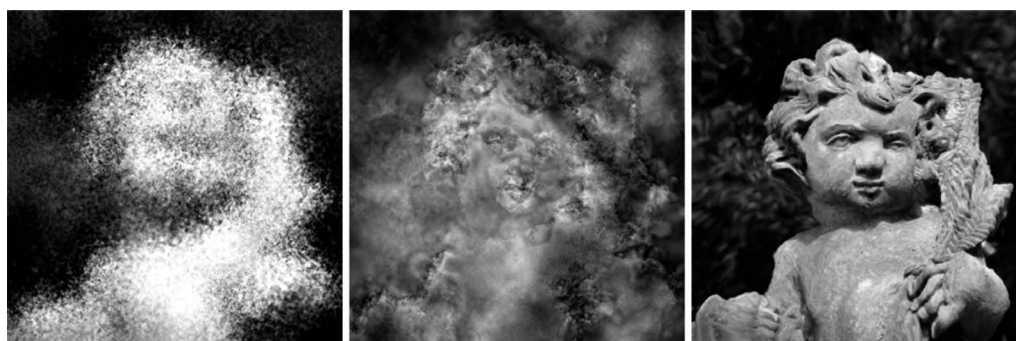


Figure 7. The effect of coherence over scales. In the image at left, all scales were independently disarrayed and equally large. In the image at center the amplitudes of dislocations were scales with the resolution. In the image at right, again scaling was used, moreover, large receptive fields dragged along smaller ones overlapped by them (on the average). In the incoherent, unscaled case (left), the dislocations are relatively most severe for the small fields, so one obtains a blurring effect. In the incoherent, scaled case, transitions (center) become diffuse and vague, whereas in the coherent case (right) they remain well defined but end up at inappropriate locations and orientations. In the latter case the result is that of a deformation. Photograph of angel statue reprinted from <https://pixabay.com/en/bust-sculpture-statue-fig-art-1555688/>.

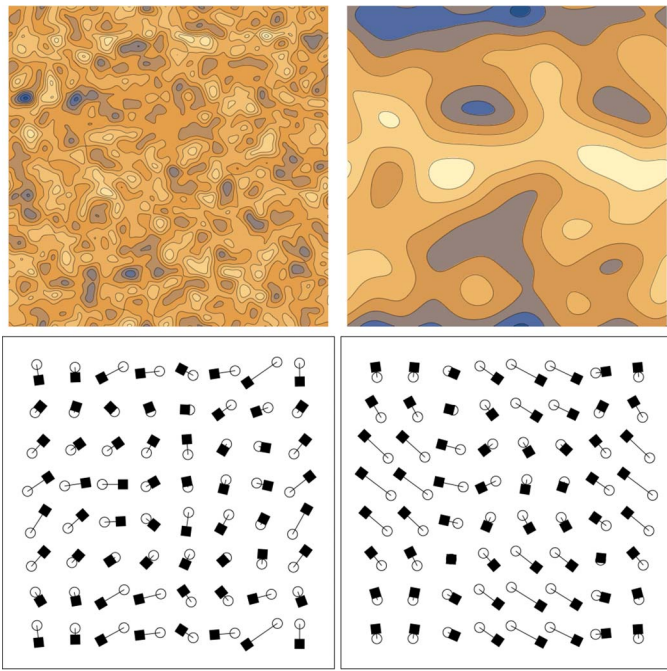


Figure 8. In the top row we show two scalar Gaussian noise fields obtained by blurring Gaussian white noise. They have different “grain,” which is parameterized by the width of the blurring kernel. In the bottom row we show disarray vector fields generated from pairs of such scalar fields. These vector fields have been severely subsampled for illustration purposes; there really resides a vector at each pixel. The columns show instances of the same grain—fine at left, coarse at right. This yields something not unlike a shuffling at left and something more like a deformation at right. Of course, the “grain” is a continuous parameter.

The example of Figure 8 already introduces one fundamental parameter, the grain. Another parameter, one that is independent of grain, we call the *reach*. The reach of the disarray (Figure 9) measures how much a

pixel will be displaced in some suitable statistical sense. Thus, it is an amplitude, or intensity-like parameter. If one scales all vectors of a random vector field by the same factor, one changes the reach, whereas the grain is not affected.

The grain and the reach suffice to parameterize many disarrays of interest. A third parameter comes into play when one regards details at different scales. We illustrate this with Figure 10.

The coherence over scale of the disarray is the third parameter. It defines the degree to which the displacement fields are correlated across scales. It takes values of zero (incoherent) if the random displacement field is generated independently for each scale; it takes a value of one (fully coherent) if the displacement fields at every scale are constructed by filtering the same Gaussian white noise samples. Coherence measures how the displacements of overlapping receptive fields of different sizes are mutually correlated. This is highly important in many applications. Coherent disarray retains the local image structure even when the global image structure is destroyed. As a result, coherent disarray appears like deformation, whereas incoherent disarray appears like diffusion or shuffling.

Each of these parameters—the grain, the reach, and the coherence—has a specific and characteristic effect. Of course, apart from the disarray, there is also the blur of the original image even before it is subjected to disarray. One may consider the blur as a fourth parameter, although it is one that is really distinct from the local sign variation. Blur and disarray can be combined in various ways.

### The implementation of disarray

How does one implement disarray? Fortunately, this turns out to be relatively simple. Methods in current use were developed for Completely Automated Public

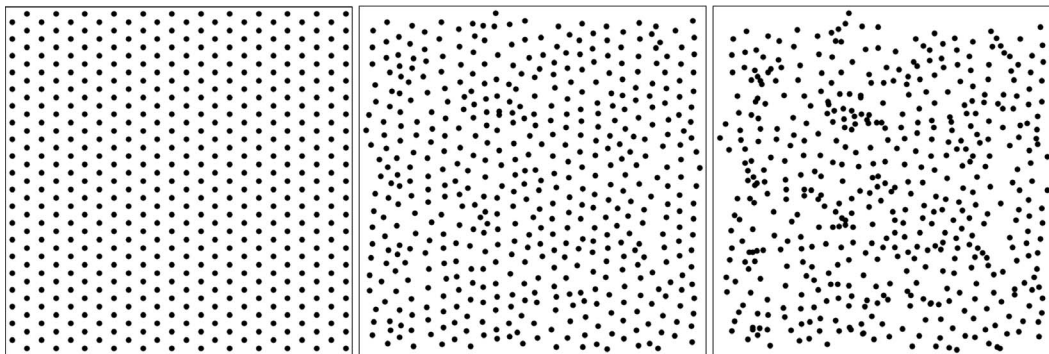


Figure 9. At left is a fiducial image—in this case, a regular hexagonal grid. If you would disarray it with a “reach” that is very small, it would not change (irrespective the grain) because all dots would remain at almost the same place. For some finite reach one obtains the eidolon shown at center. Notice that one can estimate the grain here. It is rather coarse; thus, the hexagonal structure remains locally noticeable, although it is globally destroyed. At right the image is an eidolon at much larger reach. Here the hexagonal structure is hardly retained; even neighbor relationships of the dots may have changed.

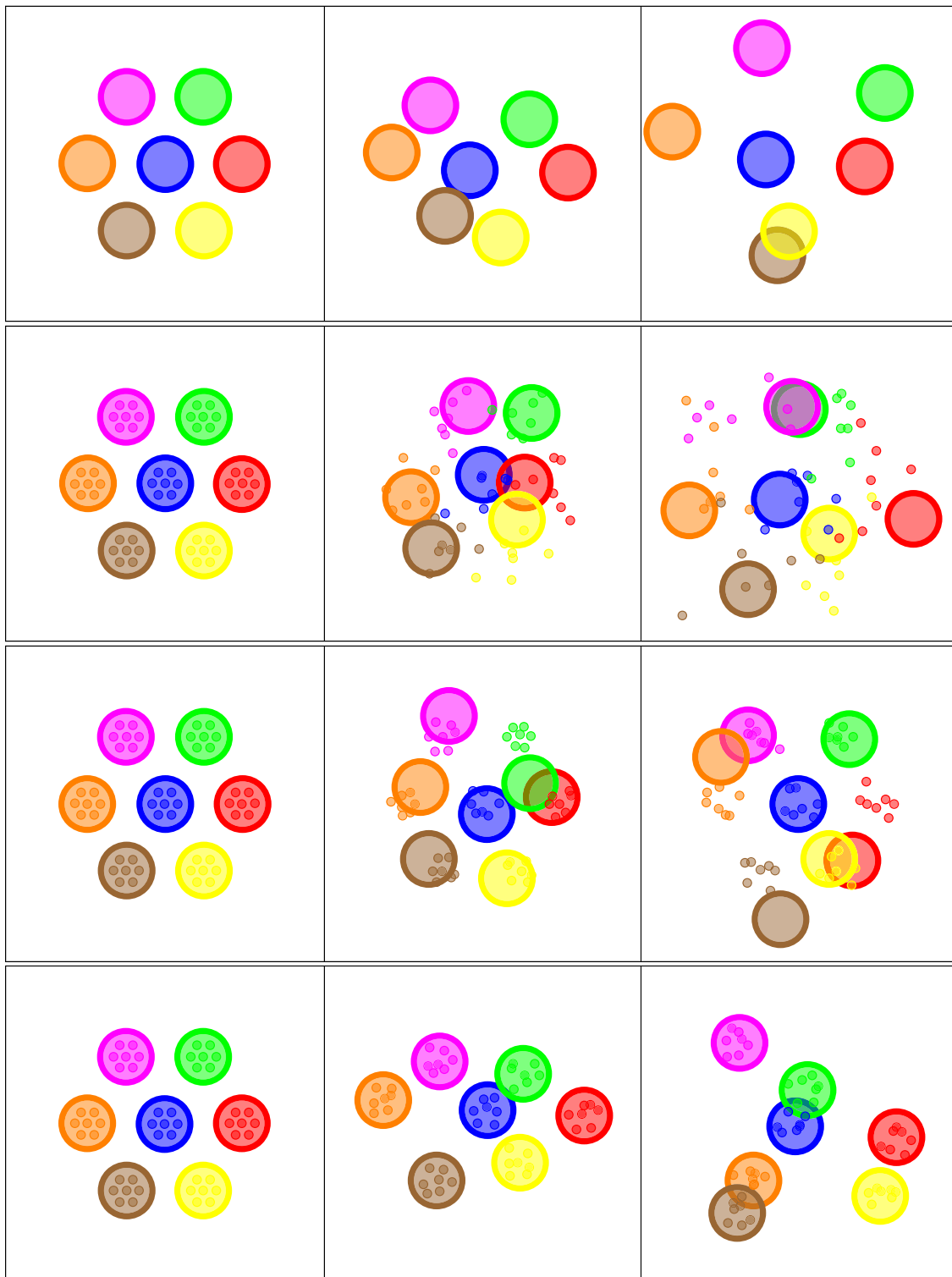


Figure 10. In the top row we show the fates of seven blobs in a hexagonal configuration under various degrees of disarray. Here we simply varied the value of the reach; the grain is such that each blob is essentially affected independently from all others. In the second row we added 49 ( $7 \times 7$ ) small blobs, each large blob containing seven small ones. The grain is such that each blob, large or small, is essentially affected independently from all others. We apply the same reach to all blobs. The effect can be seen in Figure 7 left. In the third row we treated the small blobs to a proportionally small reach compared with the large blobs. Notice how they stay together but lose their relation to the large blobs. The effect can be seen in Figure 7 center. In the bottom row the large blobs drag along the small blobs, although the latter are also individually displaced. In this case the inclusion relations are mostly retained; this is “coherent” disarray. The effect can be seen in Figure 7 right.



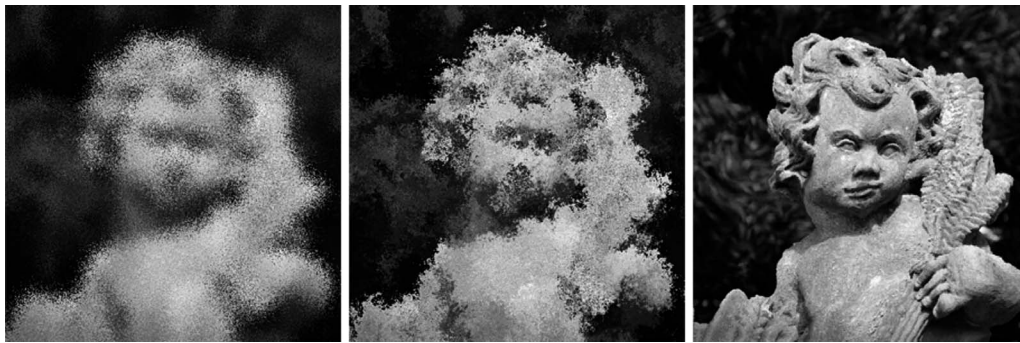


Figure 11. These are results from fractal noise in which the fractal exponent, which controls the slope of the spatial power spectrum, favors either the finer or the coarser scales. This parameter yields a convenient and intuitive handle on the style of the resulting eidolons. From left to right one has a case of white noise (flat spectrum), of pink noise and of Brownian noise for the spatial power spectrum of random displacements. Notice that far greater disarray is possible; we do not illustrate it because it soon yields fully unrecognizable images. Such images are often interesting from an artistic perspective, though (Gombrich, 1963). Photograph of angel statue reprinted from <https://pixabay.com/en/bust-sculpture-statue-fig-art-1555688/>.

Turing Test to Tell Computers and Humans Apart (CAPTCHA; von Ahn, Maurer, McMillen, Abraham, & Blum, 2008) or the automation of XKCD style graphics (Torres-Manzanera, in press).

The eidolon factory applies disarray to all elements, such as the DOG layers, that are summed over in the synthesis stage. Thus, an important issue involves the distribution of disarray over the layers and the possible interlayer correlations.

How local sign might be uncertain depends on its physiological causes, which are largely unknown. There exist essentially four mutually distinct theories: (a) Lotze's original notion that local sign is due to the experienced effects of eye movements; (b) Helmholtz's insight that local sign might be due to correlation between neural signals, the mind interpreting correlation as a sign of spatial overlap in the sensitive body surface (Koenderink, 1984a); Platt's (1960) notion that retinal translation due to eye movements might reveal sets of receptive fields in mutually collinear positions; and (d) Ahissar and Arieli's (2001) notion that "space might be coded by time" in analogy to the vibrissae systems present in most mammals (e.g., think of the whiskers of cats, rats, and so on).

One would expect very different effects in these cases. In the eidolon factory we keep an open mind and consider all such propositions as possibly simultaneously effective. This implies that a variety of disarray styles should be considered. This yields an important handle on the potential results.

Important distinctions appear to be the following. First, all receptive fields are disarrayed by the same (statistical) amount regardless of size (Figure 10, row 2). One expects such a disarray style in case the Lotze mechanism was dominant. Second, the reaches are proportional to the receptive field sizes (Figure 10, row 3). One expects such a disarray style in case the Helmholtz mechanism was dominant. Third, the

displacements of overlapping small and large receptive fields are mutually independent (Figure 10, rows 2 and 3 are examples). This might conceivably occur—in different ways—for any mechanism. Finally, when large fields are displaced they drag the small fields they overlap along with them (Figure 10, bottom row). One expects this especially in case the Platt or Ahissar mechanisms are in play.

In many cases, we need to build stacks of displacement fields of various grain sizes. In some cases, these need to be postprocessed in order to introduce correlation. An important example would be the fractal disarray, where the grains are proportional to receptive field scale and large fields drag smaller ones with them (Figure 11).

Of special interest is the layer at the finest scale. (Notice that this also contains the contributions of all coarser scales.) It is a fractal disarray field that might be used to perturb the fiducial image. This is perhaps the simplest eidolon; it simulates the tarachopic condition described by Hess (1982).

## The synthesis of eidolons

In this article we do not so much present a fixed algorithm as a class of intimately related methods. Instead of a single algorithm with several (perhaps many) parameters, we suggest a *toolbox* in which one selects the most appropriate method or combination of methods for any problem to implement certain desiderata in the simplest and most efficient way.

The toolbox method lets one zoom in in an intuitive manner, with the nature of the remaining parameters being immediately apparent. This allows you to focus on the phenomena themselves. No doubt most eidolon factory implementations will be ephemeral, constructed

for a specific purpose. However, the toolbox will be fixed; at most it will be added to, although only sparingly.

Related to this is that we do not consider the eidolon factory to be a straight emulation of the cortex at all. What we try to provide is an interface to the phenomenal complexity, whether in the mind, the brain, or the world. Good interfaces hide irrelevant complexity and are therefore very different from emulations (Hoffman, 2009; Koenderink, 2011). The toolbox preferably implements summary accounts rather than simulations of complexities that are irrelevant to the task at hand. Ideally, the eidolon factory should be trivial and thus conceptually transparent.

Here are some examples of this type of approach. First-order directional simple cells are similar to edge finders, or edge detectors. In our formalism, they are tangent vectors at some specific scale. Because the tangent space is two dimensional, a basis of just two of these—most conveniently with orthogonal direction preferences—suffices at each location.<sup>13</sup> This is functionally equivalent to the overdetermined continuous basis in the cortical implementation, for any directional sensitivity can be implemented on the fly (often known as *steerable filters*; W. T. Freeman & Adelson, 1991). This is a natural consequence of formal differential geometry (Koenderink & van Doorn, 1992). Of course, in real life one needs to consider the effects of perturbations. For instance, removing one basis vector hardly matters to the cortical overdetermined basis but would render the simple formal system inoperative.

Something similar goes for the second-order structure. In the cortex this is the system of line finder simple cells. In the formal treatment, one requires a basis of three items: either components of the Hessian (a symmetric tensor) or second-order directional derivatives at 60° orientation increments. Thus, a cortical column can be summarized by such a triple. Again, any orientation can be implemented on the fly. There is no loss of computational possibilities at the cost of being a mere summary account.

Starting with these differential operators of order less than three, there are a number of relations that are of immediate importance to possible synthesis and thus the eidolon factory. The addition of line finders at all orientations at a given location yields the Laplacian operator, which is the DOG profile for infinitesimal (small) size difference. Indeed, the DOG layers give a visually intuitive edge representation at a given scale. This makes intuitive sense because the difference of sharp and blurred instances of an image retains just the boundary regions. The edges are drawn as the Pinna watercolor illusion double lines (Pinna, 1987, 2008; Pinna et al., 2001, 2003). Because all DOG layers add up to the fiducial image, one sees that the image can be

regarded as synthesized from its edgelets. This is indeed a formal theorem (Koenderink et al., 2015).

With such relations in mind, it is understandable that perturbations in a complete cortexlike emulation can be captured by summary methods at various levels, from that of the simple cells all the way to the fiducial image. When this is indeed possible, one should opt for it because it allows one to ignore much detail that is actually causally ineffective.

This is the very notion behind the proposal of the eidolon factory as a toolbox. In psychophysical experiments it would make good sense to start with the simplest eidolons and progress to more complicated instances—perhaps eventually going all the way to the simple cells—when the empirical data cannot be described in the simpler way. After all, understanding vision means understanding it *conceptually*, not building an emulation of brain events that can itself hardly be understood because of an overdose of causally ineffective complexity.<sup>14</sup>

### Single-scale disarray

There are infinitely many ways to define eidolons. JPEG file formats and popular applications such as Adobe's Photoshop build on that. Any printer in default mode will deliver another instance. And so forth. However, we'll let that be.

In the simplest case, we skip scale space and just stay with the fiducial image. Of course, there is also an eidolon that involves scale; all renderings that cannot be distinguished because of your limited visual acuity are instances of that. It depends on whether you can find your spectacles. Here we consider only disarray at a scale level that you can easily resolve.

The resulting image is both blurred and locally scrambled (Figure 12). It will easily account for various types of amblyopia and other visual defects. It should probably be the rock-bottom start in most psychophysical investigations.

Notice that the scale is not really a parameter but more like an initial choice. There are only two essential parameters because these could be specified in terms of scale as a unit. Also notice that, when applied to unblurred images, the algorithm is effectively equivalent to the band-pass noise image distortion used by Bex (2010).

### Scale-dependent disarray

The visual field is roughly scale invariant. Not being dedicated to any specific scale rules out the single-scale disarray discussed above. One needs to acknowledge the existence of many scales. This cannot be done by mere blurring. Fourier-based methods cannot deal with this. The method somehow has to recognize the

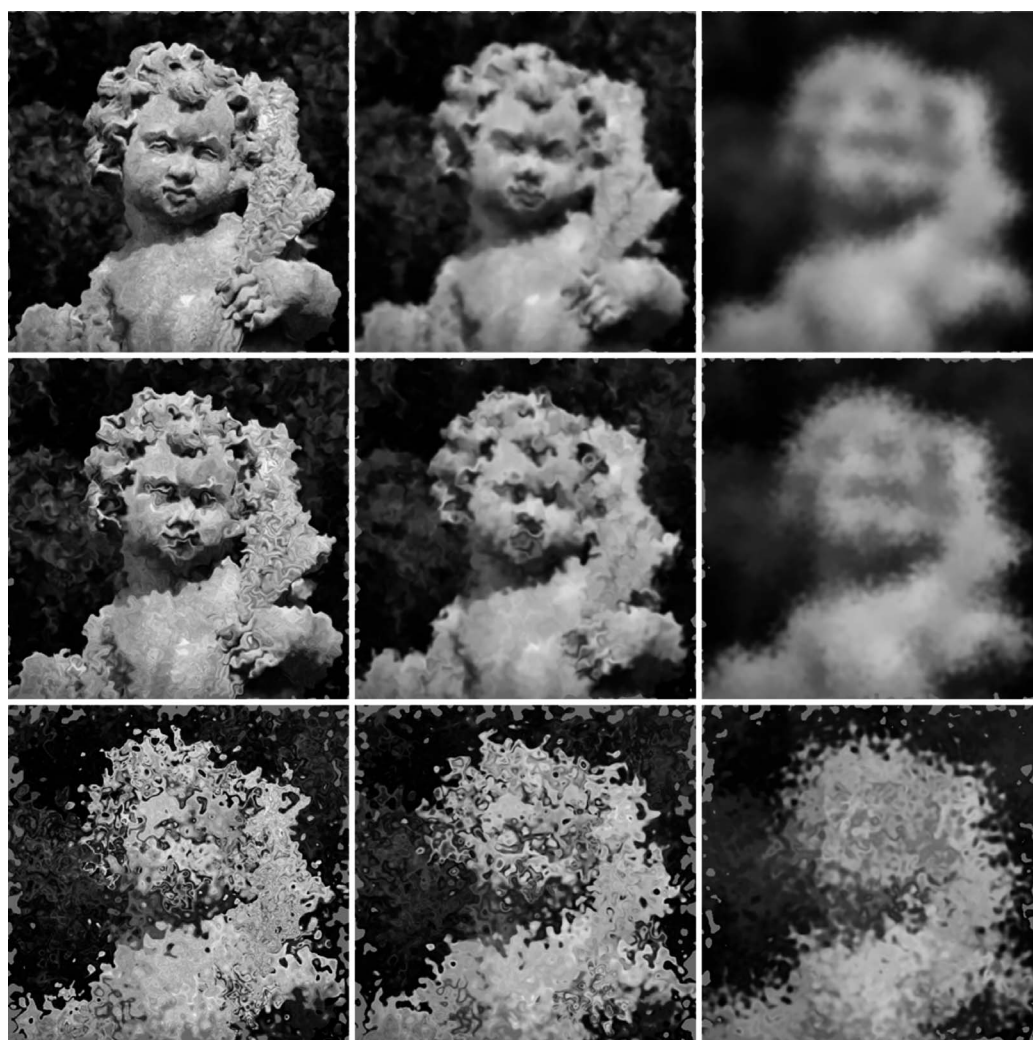


Figure 12. The combined effect of blur and disarray in various proportions. Each column is at a fixed scale of blurring; each row is a fixed reach. Notice that the blurred versions are apparently improved by disarray, although the effective resolution actually gets worse. It is simply more pleasant to look at. Blurred pictures look unsharp and are hard to focus on, which is probably why people dislike them. It is why grainy photographs of the 1960s look sharper than modern electronic ones even when the effective resolution is actually less. Photograph of angel statue reprinted from <https://pixabay.com/en/bust-sculpture-statue-fig-art-1555688/>.

spectrum of scales. The simplest method that does this is fractal disarray. It simply disorganizes the fiducial image, but it does this in a scale-independent manner.

There are numerous ways in which one might implement scale-dependent disarray. We implemented a very simple parameterization for demonstration purposes. One simply distributes the degree to which large fields drag smaller ones with them monotonically over the scale domain—that is, the degree to which the random displacement at any given location is correlated across scales. The efficaciousness of this relation then becomes the control.

This control effectively controls the fractal dimension of the local sign displacement field. It has a huge influence on pictorial structure. This is the coherence. It is a parameter of considerable conceptual interest and,

in our view, a major aspect of human vision that is hardly documented in the standard textbooks.

### ***Selection at the basis levels***

Instead of simply summing over scale-space layers (or DOG activity), one might go a stratum deeper and pool over line finders. When you apply local disarray to simple cell activity, this has little effect beyond what you obtain from the DOG activity when the reach is not too large. You will mainly notice a loss of contrast. For larger reaches you notice that edges split into several mutually independently dislocated copies, yielding an impression of superimposed ghost images (Figure 13). When the reach is very large, you obtain a mess that is hard to distinguish from what you get from the DOG activity.





Figure 13. These images are due to the coherent disarray of line finder activity. Here we use a representation with the minimum basis of three orientations, mutually separated by  $120^\circ$ . This shows up in the medium reaches, where one distinguishes ghost images. The primary visual cortex uses an overcomplete, continuous basis. In that case the image would merely grow diffuse as the superposition of arbitrarily many ghost images (of course, the precise structure depends critically on the statistical nature of the disarray too!). For these examples the reach was varied by a factor of two from instance to instance. The largest reach (bottom right) yields an image that might as well have been obtained from the DOG representation. Photograph of angel statue reprinted from <https://pixabay.com/en/bust-sculpture-statue-fig-art-1555688/>.

When you randomly leave out line finder activity, you mainly lose contrast and gain some speckle noise (Figure 14, left). Things are more complicated when you select in a focused manner. For instance, you may select line finder activity (which represents local transitions) based on edge finder activities (which signal where transitions are located). As perhaps to be expected, you obtain a drawing of the main features (Figure 14, right). Although resolution is good, some structures are simply omitted. You also lose gradual transitions.

By focusing on the basis of differential invariants (curvatures or corners, say; Attneave, 1954), you may implement many different styles of representation. Again, using the toolbox approach will easily allow you to zoom in on the structures of your choice.

## Examples

The proof of the pudding is in the eating. In the next section we present a few experiments that could be

conducted using stimuli produced with our eidolon factory. In the first, dealing with the simulation of tarachopic amblyopia symptoms in healthy observers, we show preliminary data. For the other examples, we just detail the methods and the stimuli that could be used.

## Testing tarachopic vision

Here we illustrate the use of eidolons to simulate deficits akin to the ones observed in the amblyopic condition of tarachopia in generic observers. Tarachopia is a visual agnosia described by Hess in 1982. Hess described a form of amblyopia in which an observer had two perfectly good eyes according to standard criteria. That is, the MTF functions for sine-wave grating contrast thresholds were virtually identical for both eyes, with good contrast detection threshold and good high spatial frequency roll-off. Yet, in the visual awareness of the observer, one eye was normal, whereas the other was unfit to read the headlines a newspaper. It was classified as amblyopic, which is why the patient had sought treatment. So what was wrong?



Figure 14. At left is a synthesis in which 85% of the simple cell activity was randomly deleted. The result is mainly a loss of contrast; details remain well preserved. At right, 15% of the line finder activity was kept based on the magnitude of the local r.m.s. edge finder activity. This yields an abstraction—a drawing of the features. Resolution is good, but much is left out and most smooth gradation is lost. This is the type of result that is out of reach of global methods like Fourier-based filtering. Photograph of angel statue reprinted from <https://pixabay.com/en/bust-sculpture-statue-fig-art-1555688/>.

Hess had the bold idea to let the patient report on the nature of immediate visual awareness. Compared with the “good” eye, the visual awareness of the “bad” eye appeared spatially scrambled to the patient. Hence Hess’ term *tarachopia* for this form of amblyopia, which means “scrambled vision.”

Tarachopia is a paradigmatic case because it shows the categorical difference between psychophysics proper and experimental phenomenology (Albertazzi, 2013; Koenderink, 2015a). In psychophysics proper, one uses only objective methods. But awareness is about subjective facts, a topic of phenomenology. When Hess measured contrast detection thresholds for sine-wave gratings, he was practicing psychophysics. But when he asked the patients what they saw, he switched over to experimental phenomenology, where objectivity is replaced with intersubjectivity.

Thus, the topic is a conceptually highly interesting one with important potential implications for neuroscience, psychophysics, and phenomenology. This is why we propose to study tarachopia in generic observers by imposing controlled spatial disarray on the stimulus.

We attempt to create artificial tarachopia in normal observers using eidolons. This yields a way to the classification of possibly distinct forms of tarachopia that can potentially be explored through variation of parameterization of the eidolon factory. It thus opens up a novel field of endeavor.

We depart from the standard MTF analysis of spatial vision (Cornsweet, 1970; Van Nes & Bouman, 1967), the detection thresholds for sine-wave grating modulations of an otherwise uniformly gray field. Generic human observers fail to detect gratings at spatial frequencies higher than about 50 cycles/degree and detect a range of intermediate spatial frequencies at Michelson modulations somewhat below 1%. Such analysis was introduced in television engineering by Otto Schade (1948, 1956), who measured the first MTF.

One typically records the detection of a grating as compared to a uniform field. A more objective method records the ability of observers to discriminate between horizontal and vertical gratings. For generic observers, these methods yield very similar results. However, in the case of tarachopic amblyopes, the difference is categorical. Such observers detect the presence of gratings just as well as any other, yet they fail in the ability to discriminate the orientation. They detect a pattern but fail to identify it. Hess’ suggestion that this might be due to their scrambled visual field makes intuitive sense.

In order to study this suggestion in generic observers, we produce eidolons of sine-wave gratings using a coherent local sign disarray. Representative examples of stimuli are shown in Figure 15.

The technical details of the experiment are pretty standard. The field of view was  $10^\circ \times 10^\circ$ , the average luminance of the uniform surround ( $30^\circ \times 20^\circ$ ) was 400  $\text{cd/m}^2$ , the viewing distance was 57 cm, and the

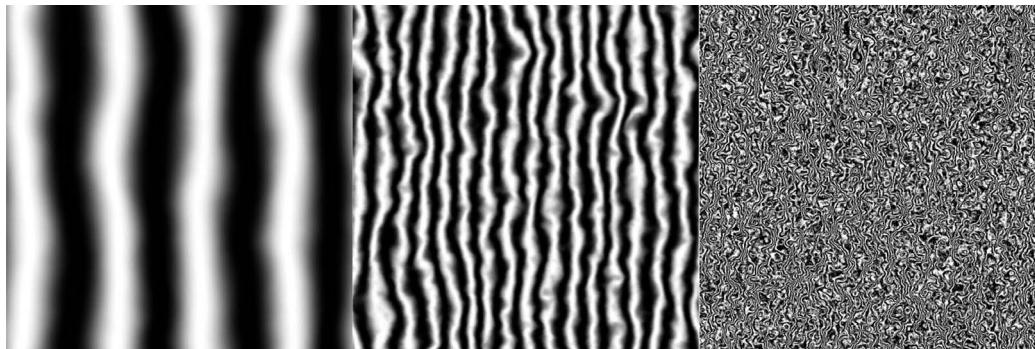


Figure 15. Examples of instances of eidolons of sine-wave gratings (here with vertical bars) of various spatial frequencies (quoted in cycles per degree). Especially at the high spatial frequencies, they look rather different from generic sine-wave gratings.

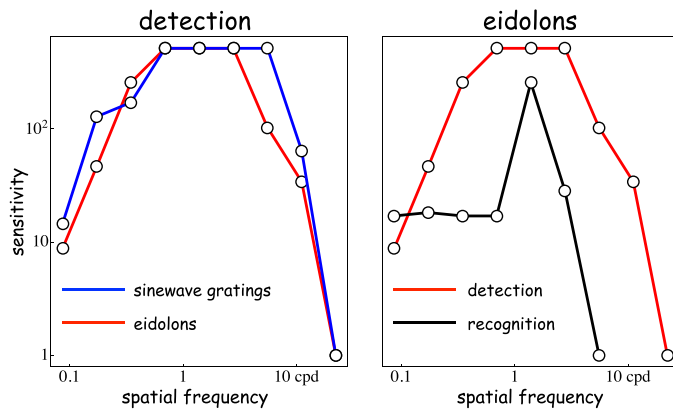


Figure 16. At left are the MTFs for detection of intensity modulations of any kind. The blue curve is for sine-wave gratings; the red curve for the eidolons. At right is the MTF for detection of eidolons in red (same as the red curve on the left) compared with the MTF for recognition of the orientation of the grating bars of the eidolons (black curve). For true sine-wave gratings, observers are aware of the orientation of the grating bars at threshold; thus, the curves for detection and recognition coincide (blue curve at left). The spatial frequency in cycles per degree, the “sensitivity” is defined as the reciprocal of the Michelson threshold contrast; this is the conventional plot.

experiment was conducted in a dark room. We used a smooth transition of the grating to the uniform field of 2° wide. The display was linearized. We used an 8-bit digital analog converter, which is only marginally suited to the usual MTF measurements but yields ample resolution for the eidolons. Observers were the authors. They have good spatial vision and fixated the center of the screen. Notice that there is no need for observers to be naïve in threshold experiments.

We performed two measurements. In the first we determined the MTF function for seeing “something” instead of “nothing.” In the second we determined the threshold for discrimination between horizontal and vertical gratings. In the first method we aimed at the 50% detection threshold, and in the second we aimed at the 75% discrimination threshold. Thresholds were determined by way of simple up–down methods. Results are not different for the three observers; we show the overall average in Figure 16.

For the detection of contrast, irrespective of pattern, there is no appreciable difference between the sine-wave gratings and the eidolons. Both contrast sensitivity and high spatial frequency roll-off are normal. But the case of recognition—in this case, of the pattern overall impression of horizontal or vertical orientation—the eidolons suffer badly.

In the case of the true gratings, observers are aware of the orientation at threshold. But in the case of the eidolons, observers may have a hard time making out

the orientation even if they see the contrast modulations well enough. Of course, this is hardly a surprise given the nature of the stimuli (Figure 16)! The paradigm put the generic observer in a position that is similar to that in which the tarachopic amblyope finds herself when viewing pure grating patterns.

Notice that the eidolon paradigm neatly emulates Hess’ findings for a tarachopic observer in generic vision. To the extent that all psychophysical testing yields the same results, the subjective report of the amblyope can be replaced with the stimulus description (Figure 15). Thus, in a certain sense, the eidolon paradigm yields an objective emulation of the subjective report. Of course, the direct proof of perceptual equivalence could be achieved only when tarachopic observers are confronted with intact images in the affected eye and eidolons in the healthy eye. We suggest this as an obvious development of this line of research.

## Matching peripheral vision in central vision

As we anticipated before, one way of measuring the difference in appearance between central and peripheral vision is to have observers draw what they see in the periphery (e.g., Metzger, 1936). Even proficient artists, however, can be expected to be able to draw reliably only relatively simple patterns; for more complex stimuli this is not a viable technique. The other option is to have observers judge the equivalence of stimuli in the center and in the periphery while changing them along some parameter. This is the approach that has been used effectively by Galvin, O’Shea, Squire, and Govan (1997) to show that peripheral stimuli can appear sharper than they actually are. We suggest that a similar approach could be used varying stimuli along the parameters of our eidolon factory and that this could provide insight into the way our visual system constructs peripheral appearance (Figure 17).

## Perception of gloss

One important question in the domain of visual perception of material properties is whether the perception of gloss depends on the global configuration of an image—for example, the congruency of highlights with the three-dimensional (3D) interpretation of the scene (Anderson & Kim, 2009; Kim, Marlow, & Anderson, 2011). The eidolon factory allows for the creation of stimuli in which the global properties of the image are destroyed by applying a long-reach disarray. At the same time, the local cross-scale structure of the image can be retained by keeping the coherence value



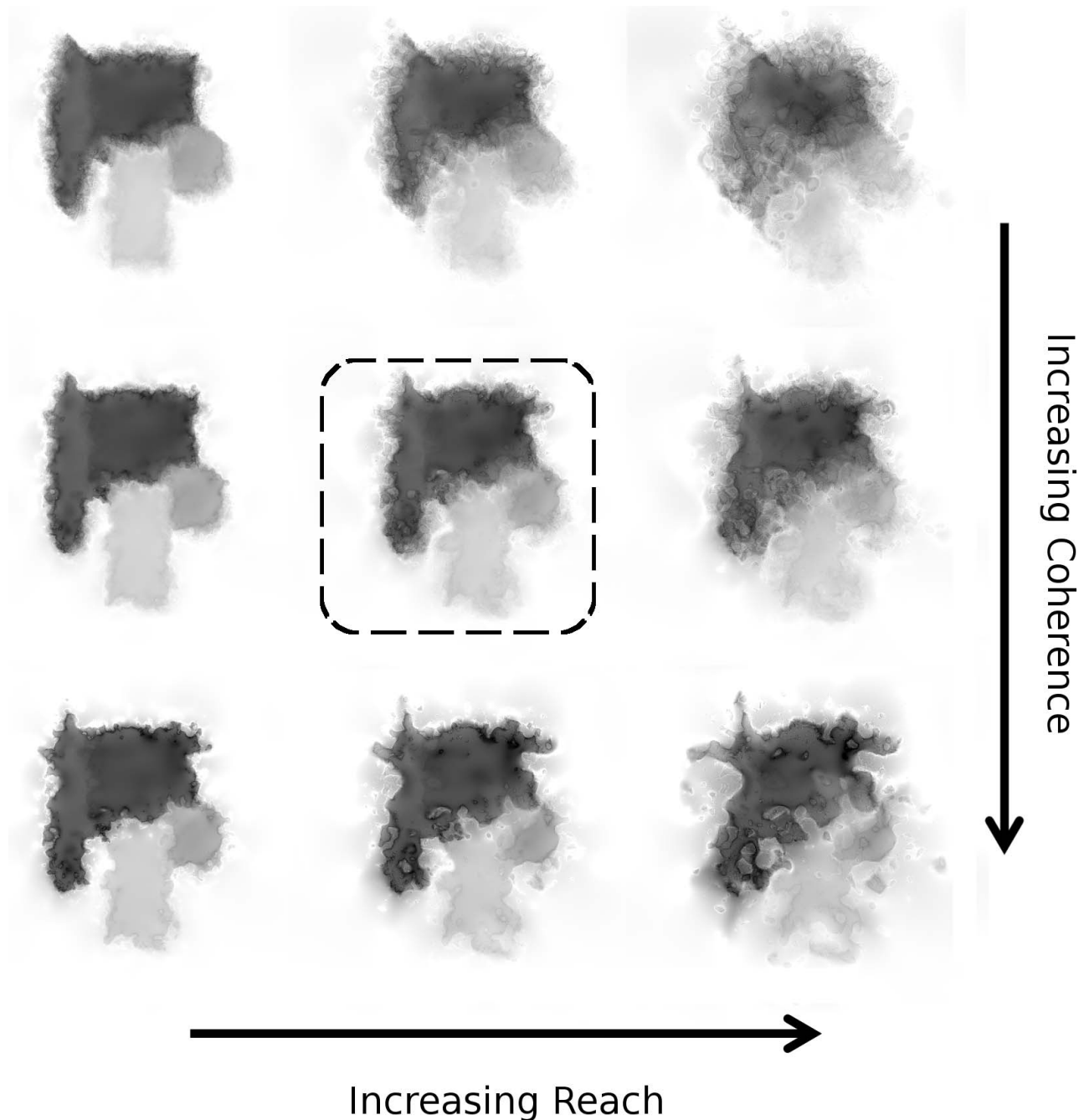


Figure 17. Examples of eidolons with different reach and coherence. Observers could be shown in peripheral viewing the central pattern and asked to navigate through the eidolons space in order to find a perceptual match in central viewing. Possible results include perfect constancy, which would be slightly boring, but finding coherence overconstancy and/or reach underconstancy could show that our visual system constructs an appearance that is more orderly than the physical input.

high. As can be seen in the images in Figure 18, it appears that the perceptual quality of gloss is to a large extent preserved when the scene structure is destroyed to the point of not being recognizable. Of course, proper psychophysical testing will have to be performed before drawing any conclusions, but this seems to suggest that local structure plays a large role in determining the perceptual quality of gloss.

### Matching touch and vision

Metzger (1936), in his laws of seeing, presented results relative to the graphical reproduction of patterns experienced haptically. Taking advantage of today's technological advances in 3D printing and image rendering, one could use the eidolon factory to test for possible distortions in the perception of

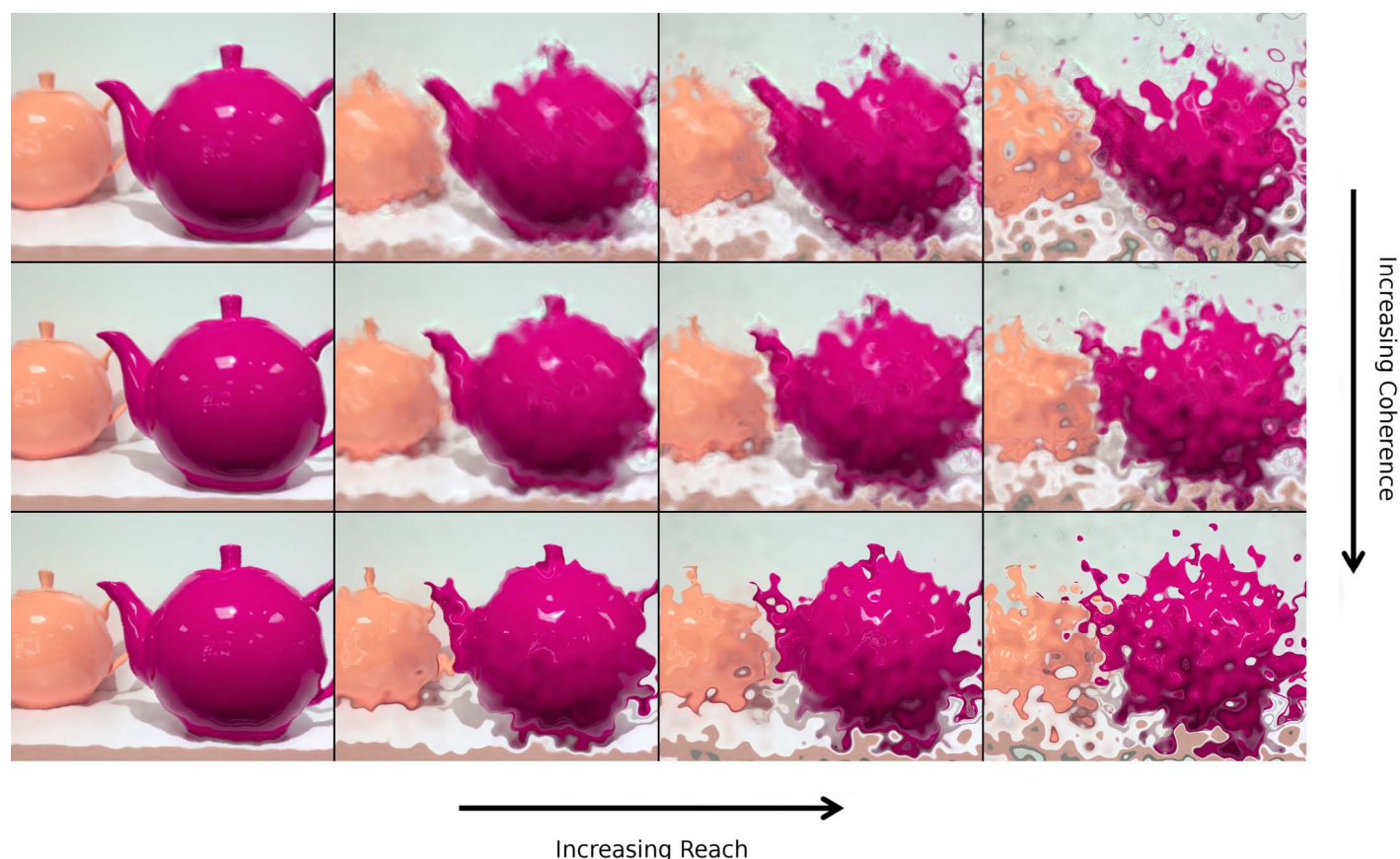


Figure 18. Eidolons of a fiducial image containing a scene with glossy objects. Notice that all figures in the bottom row (coherence = 1) look relatively glossy even when the original scene is disarrayed to the point that it is not recognizable any more. The images in the upper row (coherence = 0) appear less glossy even when the image is still recognizable.

complex haptic patterns. By taking advantage of the fact that differential geometry formalism can be applied to the local change of any quantity, we can use the eidolon factory to generate stimuli varying in relief height rather than luminance (Figure 19). Eidolons could be 3D printed for haptic exploration and rendered, possibly in 3D for visual exploration. Again, observers could be asked to navigate the parametric space of rendered visual stimuli until they find a satisfying match for the haptically experienced stimulus.

## Discussion and conclusions

There are infinitely many ways to generate eidolons. The cloud of acceptable variations on any fiducial image is huge, albeit nothing when compared with the space of all possible images. We're talking infinities here. One has to make choices. Any such choice had better be based on some fundamental considerations.

The most general methods assume nothing—neither physiological nor phenomenological nor ecological

prior knowledge. Here is a simple example, but there are numerous others. Most of the ones we can think of have been used at one or other occasion. Given a pixelated image, one simply interchanges two randomly selected pixels and repeats this a number of times. Replacing pixels with random values yields a similar result. It is technically known as “salt-and-pepper noise” (Jayaraman, Esakkirajan, & Veerakumar, 2009; Figure 20). A good parameter would be the ratio of the number of swaps to the total number of pixels. A parameter zero will return the fiducial image; a value of one will yield a totally random image. Such eidolons may well be useful in certain psychophysical contexts. From a phenomenological perspective they are trivial, and from an esthetic perspective they are appalling. They look exactly as they are—that is, alien to genesis of visual awareness. No doubt one could quantify this by computing various measures typical for cortical representations.

Other well-known examples of this general class involve adding some type of noise pattern to the image. Such methods have frequently been used in vision research. They work best with the abstract, nonsense images commonly used in psychophysics. Recognizable



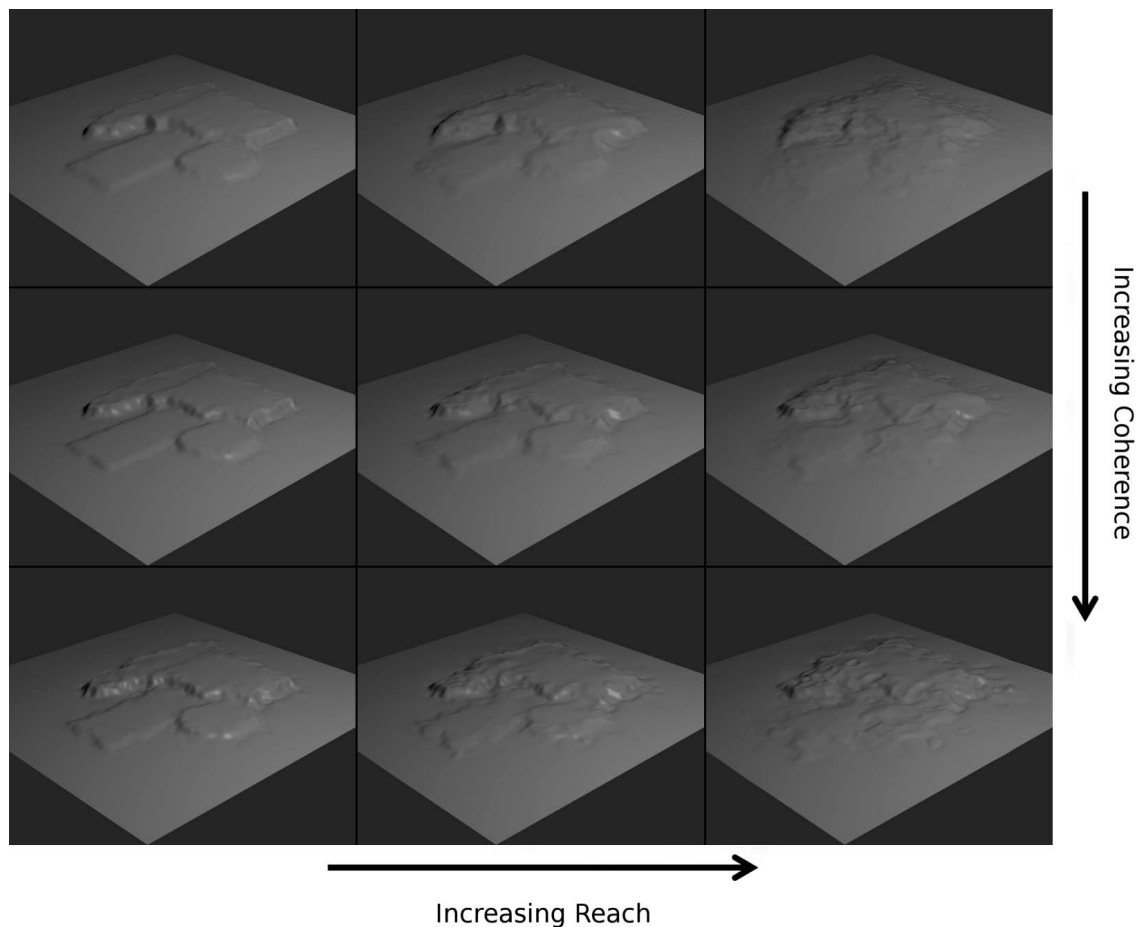


Figure 19. Renderings of possible stimuli for a haptic perception experiment. Notice that we simply mapped image intensity in the eidolon image to relief height. One could 3D print the central stimulus, have observers explore it by touch, and have the observers navigate the parametric space of rendered eidolon shapes until they find a perceptual match. We can expect that observers will tend to regularize the haptic percept, thus preferring a lower value of reach, but we are agnostic about what they will do with coherence.

images are remarkably resistant because psychogenesis is expert at beating the “cocktail party effect” (Bronkhorst, 2000; Shinn-Cunningham, 2008), which is why the familiar signal-noise methods from engineering (Kailath, Sayed, & Hassibi, 2000; Kay, 1993; Scharf, 1991) are unlikely to apply.

Other methods base eidolon generation on the essentially arbitrary way digital images are conventionally stored. A case that has become famous in vision research uses blocking (Harmon, 1973; Harmon & Julesz, 1973). Such eidolons have nothing to do with the intrinsic structure of images (e.g., why not have a

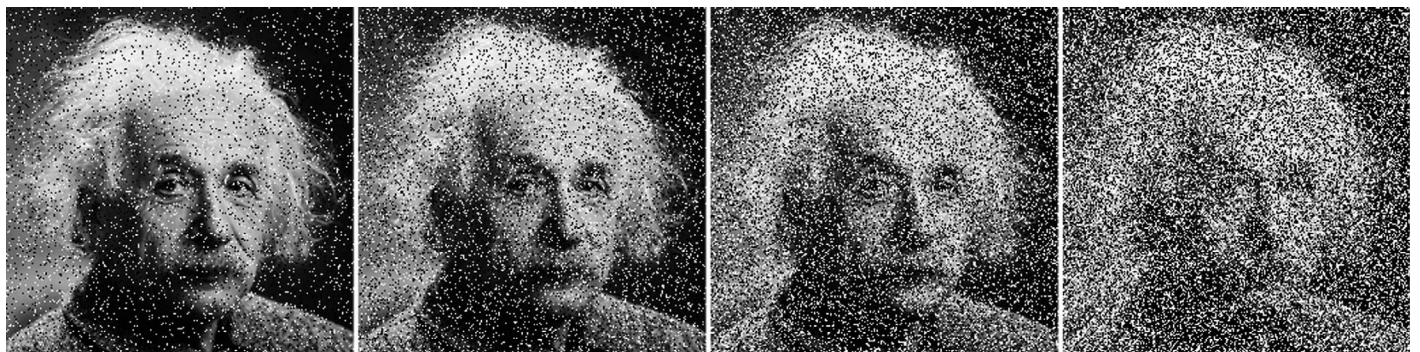


Figure 20. Examples of eidolons generated by randomly interchanging various fractions of the pixels. This is often called “salt-and-pepper noise.” It looks indeed much like noise that can fairly easily be ignored, whereas the image shines through more or less intact. Photograph of Albert Einstein reprinted from <http://www.loc.gov/pictures/item/2004671908/>. Copyright Orren Jack Turner.





Figure 21. Perhaps we should call these “Harmon and Julesz eidolons” (Harmon, 1973; Harmon & Julesz, 1973). These are interesting. Notice that a huge amount of image information is discarded, yet the gist remains visible if you look through your eyelashes, as painters do, or add some noise or apply blur to mask the sharp edges, as vision researchers do. The pixel subsampling is entirely unrelated to image content, neurophysiology, or phenomenology. Photograph of Albert Einstein reprinted from <http://www.loc.gov/pictures/item/2004671908/>. Copyright Orren Jack Turner.

honeycomb array of hexagonal pixels instead of a Cartesian checkerboard?), the known physiology, or the phenomenology of the visual field (Figure 21). They allow easy parameterization of structural complexity and are inherently local—both desirable properties.

Well-known instances are phase-scrambled images (Oppenheim & Lim, 1981; Thomson, 1999; Vogels, 1999). Here, disarray is applied in the spectral domain. The methods were perhaps inspired by the notion that sine-wave gratings are natural parts of images or somehow special to what the primary visual cortex is up to. Both (mutually related) notions are false, but that is not the point here. What makes Fourier analysis special is that it is global. This has indeed some—though not much—relation to what might be desirable for a biologically viable optic sensor system. The eidolons one obtains look somewhat better than those from the previous example (Figure 22), yet they don’t look natural. Apparently, the *global* nature of the parts is problematic.

Engineers do what is possible and most economical, science not being their first priority. Yet, historically, engineers have produced eidolons of considerable interest and importance. We simply mention a few—there are many—of their achievements. Early television engineering was much about bandwidth. Otto Schade (1948) pioneered the MTF, enabling him to create eidolons that were acceptable to the public for many years. The development of color television yields a similar story. Eidolons were based on opponent channels, with the higher bandwidth devoted to the luminance channel. When digital images became common, the aforementioned JPEG eidolons became of major importance. They are based on rather intricate properties of spatial vision.

Of course, it is much more interesting to consider eidolons that are somehow constrained by scientific understanding of either the neurophysiology or the phenomenology of visual awareness (Stojanoski & Cusack, 2014). A well-known current model of eidolons is due to the work at Eero Simoncelli’s lab at New York

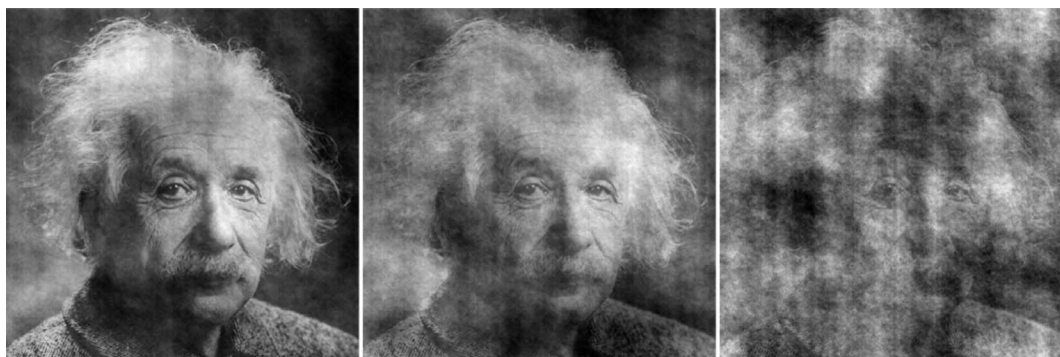


Figure 22. These are phase-scrambled images. The (mainly) vertical linear artifacts are due to the edges of the image, the right side being darker than the left side. (Simple Fourier analysis/synthesis treats the image as periodic in the horizontal and vertical dimensions, thus the left and right side compose a sharp edge of considerable contrast.) Photograph of Albert Einstein reprinted from <http://www.loc.gov/pictures/item/2004671908/>. Copyright Orren Jack Turner.



Figure 23. In this example we used space-variant disarray to place emphasis on either the chimpanzee on the left or on the right. The possibility to very easily modulate disarray spatially is a great advantage of the eidolons proposed by us, as opposed to most contenders. Photograph of chimpanzees reprinted from <https://pixabay.com/en/monkeys-chimpanzees-savages-group-1200216/>.

University (Portilla & Simoncelli, 2000). Their idea is to impose empirical knowledge of the structural complexity of neural activity in V1 as a constraint on the structure of the eidolons. Thus, an eidolon is an image that would ideally evoke the statistically equivalent neural activity as the fiducial image does. This is an important notion, characterizing the bottleneck imposed by V1 (or anything up to V1), very similar to—but hugely more complicated than—the notion of metamerism in colorimetry.

Rather than just analyzing the fiducial image first and then trying random images until you hit on one that has the equivalent statistics, the actual implementation forces an initial noise image into complying with the descriptor values established in the analysis stage. Otherwise, it would be like waiting for the proverbial monkey randomly hitting the keys of a typewriter to finish a perfect transcript of Helmholtz's *Handbook of Physiological Optics*. It is a sculpting of essentially random structures to be as equivalent to the fiducial statistics as can be. We describe these methods in Appendix C.

The Portilla-Simoncelli algorithm produces the “mongrels” that have been used in the highly interesting research in Ruth Rosenholtz's lab at Massachusetts Institute of Technology (Rosenholtz, 2011; Rosenholtz, Huang, & Ehinger, 2012; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012). They have pioneered the use of mongrels as a novel and powerful tool in vision research. That is exactly the intended use of the eidolons proposed here.

The eidolons are based on the phenomenology of vision rather than neurophysiology. However, on the formal level there are obvious tangencies. Our inspiration came also from the study of painting methods. Visual artists, throughout the centuries, have been involved in the production of eidolons. On the whole they have been very successful, their clients sometimes

taking their eidolons for reality. However, they explored the territory thoroughly and reached the dark boundary regions where the eidolons fall apart, perhaps taxing the visual competence of some observers but leaving the majority of their public behind. We find that technical painting methods are closely related to the phenomenology of vision as studied academically (e.g., Cateura, 1995; Jacobs, 1986). Our eidolon factory is based on that.

The eidolon factory described here (technically in Appendix B; a demonstration program is available—see Appendix A) offers some desirable features for vision research:

It is formally simple and transparent. It is essentially just the mathematician's toolbox of differential geometry (Bell, 2005; Koenderink, 1990; Koenderink & van Doorn, 1992; Spivak, 1999; ter Haar Romeny, 2008).

It is overall linear except for places where essential nonlinearities come in a transparent manner.

It is algorithmically simple and transparent. No magical numbers. No iterative procedures (Appendix C). No partial differential equations to solve (Elder & Zucker, 1998). All that happens is the accumulation of mass activity—our synthesis stage.

It is a nice summary account of what the cortex might be computing. Such summary accounts (actually caricatures, of course) might be more useful than an exhaustive description because they appeal to the intuition. Phonebooks are useful but hardly appeal to the understanding.

It is a powerful heuristic in that it is easily expandable. There are only a limited number of crucial elements to be understood.

Eidolons can be obtained in a straightforward manner; only a few (intuitively obvious) parameters need to be set.

This may sound like eidolon factories are just for squares—there being no surprises or challenges for the cool kids! But that would be too limited a picture. Being able to actually understand what is happening is really a source of freedom. The parameters at your disposal are meaningful, and their actions are largely independent of each other. So, you have an interface to the factory that is transparent. There will be no surprises once you understand the basic (simple) structure. This makes it possible to aim your investigation of visual awareness much more precisely. It puts you in control as a scientific investigator. When probing nature (including the human mind!), the surprises should be due to nature rather than the probing tool.

Because it is so simple and direct, the eidolon factory is very easily extended in various directions. For instance, one may apply disarray just as well to opponent color channels as to the orientation of edgelets and so forth. Disarray is also easy to apply in a spatially nonuniform way, opening up many directions of research. A simple example of such focused disarray is shown in Figure 23.

In conclusion, we have proposed an eidolon factory that is quite open ended in its potential applications and capable of almost endless development. It is also simple enough that one may adapt it for any specific application. Because of its simplicity, it allows one to tailor the nature of eidolons to specific problems.

*Keywords:* appearance, natural images, local sign

## Acknowledgments

KG and MV were supported by the Deutsche Forschungsgemeinschaft DFG SFB/TRR 135. MV was supported by the EU Marie Curie Initial Training Network “PRISM” (FP7—PEOPLE-2012-ITN; grant agreement 316746). JW, JK, and AvD were supported by the program by the Flemish Government (METH/14/02), awarded to JW. JK was supported by a Humboldt-Award by the Alexander-von-Humboldt-Foundation.

Commercial relationships: none.

Corresponding author: Karl R. Gegenfurtner.

Email: Karl.R.Gegenfurtner@psychol.uni-giessen.de.

Address: Justus-Liebig Universität Giessen, Abteilung Allgemeine Psychologie, Giessen, Germany.

## Footnotes

<sup>1</sup> Originally, an eidolon was a shade or phantom of a person, either living or dead. Most surprisingly,

Euripides, in his work on the Trojan War, claims that Helen of Troy was never physically present in the city but rather that the Greeks fought over such an illusion (Holmberg, 1995; Meltzer, 1994; Papi, 1987). The literal translation is quoted as “image, idol, apparition, phantom, ghost” from *eidōs* (“form”). We use it to denote a shape-shifted doppelgänger (Damas Mora, Jenner, & Eacott, 1980; Poe, 1839) of the fiducial image in a very precise sense.

<sup>2</sup> The plural *eidolons* seems better suited to technical jargon than the more proper *eidola* because North Americans are likely to prefer it. Walt Whitman’s (1891) famous poem “Eidolons” from his *Leaves of Grass* inspired our term.

<sup>3</sup> Our idea of the eidolon is also related to the concept of “duck typing” in computer science. According to Wikipedia (2016), “Duck typing is concerned with establishing the suitability of an object for some purpose. With normal typing, suitability is assumed to be determined by an object’s type only. In duck typing, an object’s suitability is determined by the presence of certain methods and properties (with appropriate meaning), rather than the actual type of the object.”

<sup>4</sup> Thus, it is useless to look for signs of the chlorophyll absorption bands in the spectrum of the radiation from the pixels that present you with a green meadow on your monitor.

<sup>5</sup> Thus, the cardinality of the set is about  $2^{10,000}$ —that is, about  $2 \times 10^{3010}$ . This is practically infinite, given that estimates of the number of fundamental particles in the observable universe go up to only  $10^{85}$ .

<sup>6</sup> The field of view is something completely different! That is simply physics, if you want.

<sup>7</sup> Of course, this can be only approximately the case because the total extent is limited (a half space in front of the observer—about  $180^\circ$ ) and so is the smallest size (visual acuity, about  $1'$ ). But it is a very meaningful approximation. All fractal structure has to fail in the very small and the very large.

<sup>8</sup> The assumptions involve the subjective notion of what an edge is.

<sup>9</sup> The Gabor functions made good sense in Gabor’s original application, which was not vision but acoustics.

<sup>10</sup> The term was generally accepted and freely used (e.g., by Helmholtz) and was a mainstream concept for the first half of the 20th century. However, it dropped out of the mainstream consciousness after (perhaps) the 1960s. A modern rendering might be something like *positional signature*. However, we prefer to hold on to the original term.

<sup>11</sup> Lotze speculated on the existence of somatotopy and cogently reasoned that it would in no way explain local sign. Suppose you manage to spatially scramble the cortex, leaving connections intact. Would that



produce a tarachopic amblyope? In the final instance it all boils down to the mind–brain issue, on which psychology has to be agnostic.

<sup>12</sup> On the concept of “extensive”: The fiducial image is one singular instance, whereas the eidolon is an extensive cloud that surrounds it, containing (potentially) infinitely many equivalent instances.

<sup>13</sup> Thus, the formal model is rather more concise than the cortex, which represents the first- and second-order structures in terms of highly overdetermined, continuous bases.

<sup>14</sup> Here is a simple example that may serve to catch the idea. No doubt the best emulation of meteorology would be the weather itself, as you would obtain perfect predictions. (Alas, its predictions would become available at the latest moment only!) But this perfect emulation would not go far in promoting the understanding of atmospheric processes. After all, it (i.e., the global atmosphere as an analog computer) was already available to Aristotle. Aerodynamics and thermodynamics yield much more insight, although the predictions are not necessarily all that great in practice.

## References

- Abramowitz, M., & Stegun, I. A. (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. New York, NY: Dover.
- Ahissar, E., & Arieli, A. (2001). Figuring space by time. *Neuron*, 32(2), 185–201.
- Albertazzi, L. (2013). *Handbook of experimental phenomenology: Visual perception of shape, space and appearance*. New York, NY: Wiley.
- Anderson, B. L., & Kim, J. (2009). Image statistics do not explain the perception of gloss and lightness. *Journal of Vision*, 9(11):10, 1–17, doi:10.1167/9.11.10. [PubMed] [Article]
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61(3), 183–193.
- Balas, B. (2006). Texture synthesis and perception: Using computational models to study texture representations in the human visual system. *Vision Research*, 46(3), 299–309.
- Balas, B., & Conlin, C. (2015). Invariant texture perception is harder with synthetic textures: Implications for models of texture processing. *Vision Research*, 115(Part B ), 271–279.
- Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, 9(12):13, 1–18, doi:10.1167/9.12.13. [PubMed] [Article]
- Bell, J. L. (2005). *The continuous and the infinitesimal in mathematics and philosophy*. Milan, Italy: Polimetris S.A.
- Bex, P. J. (2010). (In) Sensitivity to spatial distortion in natural scenes. *Journal of Vision*, 10(2):23, 1–15, doi:10.1167/10.2.23. [PubMed] [Article]
- Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects—Evidence and conditions for 3-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, 19(6), 1162–1182.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226(5241), 177–178.
- Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica United With Acustica*, 86(1), 117–128.
- Burt, P. J. (1981). Fast filter transforms for image-processing. *Computer Graphics and Image Processing*, 16(1), 20–51.
- Burt, P. J., & Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4), 532–540.
- Burton, H. E. (1945). The optics of Euclid. *Journal of the Optical Society of America*, 35(5), 357–372.
- Cateura, L. (1995). *Oil painting secrets from a master*. New York, NY: Watson & Guptill.
- Cornsweet, T. (1970). *Visual perception*. New York, NY: Academic Press.
- Damas Mora, J. M. R., Jenner, F. A., & Eacott, S. E. (1980). On heautoscopy or the phenomenon of the double: Case presentation and review of the literature. *British Journal of Medical Psychology*, 53(1), 75–83.
- DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, 11(8), 333–341.
- Elder, J. H., & Zucker, S. W. (1998). Local scale control for edge detection and blur estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(7), 699–716.
- Florack, L. M. J. (1997). *Image structure* (Vol. 10). Dordrecht, The Netherlands: Kluwer Academic.
- Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, 14(9), U1195–U1130.
- Freeman, W. T., & Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9), 891–906.

- Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers—Part III: Radio and Communication Engineering*, 93(26), 429–441.
- Galvin, S. J., O'Shea, R. P., Squire, A. M., & Govan, D. G. (1997). Sharpness overconstancy in peripheral vision. *Vision Research*, 37(15), 2035–2039.
- Gombrich, E. H. (1963). *Meditations on a hobby horse*. London, UK: Phaidon.
- Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2010). Crowding changes appearance. *Current Biology*, 20(6), 496–501.
- Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 352(1358), 1121–1127.
- Harmon, L. D. (1973). The recognition of faces. *Scientific American*, 229, 70–82.
- Harmon, L. D., & Julesz, B. (1973). Masking in visual recognition: Effects of two-dimensional filtered noise. *Science*, 180(4091), 1194–1197.
- Heideman, M. T., Johnson, D. H., & Burrus, C. S. (1985). Gauss and the history of the fast Fourier-transform. *Archive for History of Exact Sciences*, 34(3), 265–277.
- Helmholtz, H. (1892). *Physiologische Optik* (2nd ed.). Leipzig, Germany: Leopold Voss.
- Hess, R. F. (1982). Developmental sensory impairment, amblyopia or tarachopia. *Human Neurobiology*, 1, 17–29.
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5), 791–804.
- Hoffman, D. (2009). The interface theory of perception: Natural selection drives true perception to swift extinction. In S. Dickinson, M. Tarr, A. Leonardis, & B. Schiele (Eds.), *Object categorization: Computer and human vision perspectives* (pp. 148–167). Cambridge, UK: Cambridge University Press.
- Holmberg, I. E. (1995). Euripides' Helen: Most noble and most chaste. *American Journal of Philology*, 116(1), 19–42.
- Hong, H., Yamins, D. L. K., Majaj, N. J., & DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience*, 19(4), 613–622.
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London)*, 195(1), 215–243.
- Hurlbert, A. (2007). Colour constancy. *Current Biology*, 17(21), R906–R907.
- Jacobs, T. S. (1986). *Light for the artist*. New York, NY: Watson & Guptill.
- James, W. (1890). *The principles of psychology*. New York, NY: Dover.
- Jayaraman, S., Esakkirajan, S., & Veerakumar, T. (2009). *Digital image processing*. New Delhi, India: Tata McGraw Hill Educational.
- Julesz, B. (1962). Visual pattern discrimination. *IRE Transactions on Information Theory*, 8(2), 84–92.
- Kaas, J. H. (1990). Somatosensory system. In J. K. Mai & G. Paxinos (Eds.), *The human nervous system* (pp. 813–844). San Diego, CA: Academic Press.
- Kailath, T., Sayed, A. H., & Hassibi, B. (2000). *Linear estimation*. Upper Saddle River, NJ: Prentice Hall.
- Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neural science* (4th ed.). New York, NY: McGraw-Hill.
- Kandinsky, W. (1959). *Punkt und Linie zu Fläche*. Bern-Bümplitz, Switzerland: Benteli-Vlg.
- Kanizsa, G. (1997). *Grammatica del vedere. Saggi su percezione e Gestalt*. Bologna, Italy: Il Mulino.
- Kay, S. M. (1993). *Fundamentals of statistical signal processing*. Upper Saddle River, NJ: Prentice Hall.
- Kim, J., Marlow, P., & Anderson, B. L. (2011). The perception of gloss depends on highlight congruence with surface shading. *Journal of Vision*, 11(9): 4, 1–19, doi:10.1167/11.9.4. [PubMed] [Article]
- Koenderink, J. J. (1984a). The concept of local sign. In A. J. van Doorn, W. A. van de Grind, & J. J. Koenderink (Eds.), *Limits in perception* (pp. 495–547). Utrecht, The Netherlands: VNU Science Press.
- Koenderink, J. J. (1984b). The structure of images. *Biological Cybernetics*, 50(5), 363–370.
- Koenderink, J. J. (1990). The brain a geometry engine. *Psychological Research—Psychologische Forschung*, 52(2–3), 122–127.
- Koenderink, J. J. (2010). *Color for the sciences*. Cambridge, MA: MIT Press.
- Koenderink, J. J. (2011, Jan/Feb). *Vision as a user interface*. Paper presented at IS&T/SPIE Electronic Imaging 2011, San Francisco, CA.
- Koenderink, J. J. (2015a). Methodological background: Experimental phenomenology. In J. Wagemans (Ed.), *Oxford handbook of perceptual organization* (pp. 41–54). Oxford, UK: Oxford University Press.
- Koenderink, J. J. (2015b). Ontology of the mirror world. *Gestalt Theory*, 37(2), 119–140.
- Koenderink, J. J., Richards, W., & van Doorn, A. J.

- (2012a). Blow-up: A free lunch? *I-Perception*, 3(2), 141–145.
- Koenderink, J. J., Richards, W., & van Doorn, A. J. (2012b). Space-time disarray and visual awareness. *I-Perception*, 3(3), 159–165.
- Koenderink, J. J., & van Doorn, A. J. (1992). Generic neighborhood operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(6), 597–605.
- Koenderink, J. J., van Doorn, A. J., & Pinna, B. (2015). Psychogenesis of gestalt. *Gestalt Theory*, 37(3), 287–304.
- Koenderink, J. J., van Doorn, A. J., Pinna, B., & Wagemans, J. (2016). Boundaries, transitions and passages. *Art and Perception*, 4(3), 185–204.
- Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences, USA*, 96(1), 307–312.
- Lettvin, J. Y. (1976). On seeing sidelong. *The Sciences*, 16(4), 10–20.
- Levi-Civita, T., & Ricci, G. (1900). Méthodes de calcul différentiel absolu et leurs applications. *Mathematische Annalen B*, 54, 125–201.
- Lindeberg, T. (1994). Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21(2), 224–270.
- Lindeberg, T. (2013). A computational theory of visual receptive fields. *Biological Cybernetics*, 107(6), 589–635.
- Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, 19, 577–621.
- Lotze, H. (1852). *Medizinische Psychologie oder Physiologie der Seele*. Leipzig, Germany: Wedmann'sche Buchhandlung.
- Maffei, L., & Fiorentini, A. (1973). Visual-cortex as a spatial frequency analyzer. *Vision Research*, 13(7), 1255–1267.
- Margulis, D. (1998). *Sharpening with a stiletto*. Retrieved from [https://www.ledet.com/margulis/Makeready/MA27-Sharpening\\_With\\_Stiletto.pdf](https://www.ledet.com/margulis/Makeready/MA27-Sharpening_With_Stiletto.pdf)
- Margulis, D. (2005). *Life on the edge*. Retrieved from [https://www.ledet.com/margulis/Makeready/MA69-Life\\_on\\_the\\_Edge.pdf](https://www.ledet.com/margulis/Makeready/MA69-Life_on_the_Edge.pdf)
- Marr, D., & Hildreth, E. (1980). Theory of edge-detection. *Proceedings of the Royal Society B: Biological Sciences*, 207(1167), 187–217.
- Meltzer, G. S. (1994). “Where is the glory of Troy?” “Kleos” in Euripides’ “Helen.” *Classical Antiquity*, 13(2), 234–255.
- Metzger, W. (1936). *Gesetze des Sehens* (3rd ed.). Frankfurt, Germany: Verlag Waldemar Kramer.
- Monmonier, M. S. (1999). *Air apparent: How meteorologists learned to map, predict, and dramatize weather*. Chicago, IL: University of Chicago Press.
- Okazawa, G., Tajima, S., & Komatsu, H. (2015). Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proceedings of the National Academy of Sciences, USA*, 112(4), E351–E360.
- O’Neill, M. E. (in press). PCG: A family of simple fast space-efficient statistically good algorithms for random number generation. *ACM Transactions on Mathematical Software*, in press.
- Oppenheim, A. V., & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, 69(5), 529–541.
- Papi, D. G. (1987). Victors and sufferers in Euripides’ Helen. *American Journal of Philology*, 108(1), 27–40.
- Papoulis, A. (1962). *The Fourier integral and its applications*. New York, NY: McGraw-Hill.
- Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current Opinion in Neurobiology*, 18(4), 445–451.
- Pennebaker, W. B., & Mitchell, J. L. (1993). *JPEG still image data compression standard* (3rd ed.). Dordrecht, The Netherlands: Kluwer Academic.
- Pinna, B. (1987, Sep/Oct). *Un effetto di colorazione*. Paper presented at the Il laboratorio e la città. XXI Congresso degli Psicologi Italiani, Venezia, Italy.
- Pinna, B. (2008). Watercolor illusion. *Scholarpedia*, 3(1), 5352.
- Pinna, B., Brelstaff, G., & Spillmann, L. (2001). Surface color from boundaries: A new “watercolor” illusion. *Vision Research*, 41(20), 2669–2676.
- Pinna, B., Werner, J. S., & Spillmann, L. (2003). The watercolor effect: A new principle of grouping and figure-ground organization. *Vision Research*, 43(1), 43–52.
- Platt, J. R. (1960). How we see straight lines. *Scientific American*, 202(6), 121–129.
- Poe, E. A. (1839). William Wilson. *Burton’s Gentleman’s Magazine*, September, 145–152.
- Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1), 49–70.
- Reas, C., & Fry, B. (2010). *Getting started with processing*. Sebastopol, CA: O’Reilly.
- Rosenholtz, R. (2011, Jan). *What your visual system sees where you are not looking*. Paper presented at



- IS&T/SPIE Electronic Imaging 2011, San Francisco, CA.
- Rosenholtz, R., Huang, J., & Ehinger, K. A. (2012). Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology*, 3, 13.
- Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, 12(4):14, 1–17, doi:10.1167/12.4.14. [PubMed] [Article]
- Rust, N. C., & DiCarlo, J. J. (2012). Balanced increases in selectivity and tolerance produce constant sparseness along the ventral visual stream. *Journal of Neuroscience*, 32(30), 10170–10182.
- Savant, S. (2014). A review on edge detection techniques for image segmentation. *International Journal of Computer Science and Information Technologies*, 5(4), 5898–5900.
- Sayim, B., & Wagemans, J. (2013). Drawings of the visual periphery reveal appearance changes in crowding. *Perception*, 42, 229.
- Schade, O. H. (1948). Electro-optical characteristics of television systems (Television Microphotometer). *RCA Review*, 9, 527–530.
- Schade, O. H. (1956). Optical and photoelectric analog of the eye. *Journal of the Optical Society of America*, 46(9), 721–739.
- Scharf, L. L. (1991). *Statistical signal processing: Detection, estimation, and time series analysis*. Boston, MA: Addison-Wesley.
- Schmalzing, J. (1997, Oct). *Koenderink filters and the microwave background*. Paper presented at the 2nd SFB Workshop on Astro-Particle Physics, Ringberg Castle, Tegernsee, Germany.
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, 12(5), 182–186.
- Simoncelli, E. P., Freeman, W. T., Adelson, E. H., & Heeger, D. J. (1992). Shiftable multiscale transforms. *IEEE Transactions on Information Theory*, 38(2), 587–607.
- Sloan, L. L., & Wollach, L. (1948). A case of unilateral deuteranopia. *Journal of the Optical Society of America*, 38(6), 502–509.
- Spivak, M. (1999). *A comprehensive introduction to differential geometry* (3rd ed.). Boston, MA: Publish or Perish.
- Stojanoski, B., & Cusack, R. (2014). Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. *Journal of Vision*, 14(12):6, 1–16, doi:10.1167/14.12.6. [PubMed] [Article]
- ter Haar Romeny, B. M. (2008). *Front-end vision and multi-scale image analysis*. New York, NY: Springer.
- Thomson, M. G. A. (1999). Visual coding and the phase structure of natural scenes. *Network: Computation in Neural Systems*, 10(2), 123–132.
- Titchener, E. B. (1902). *Experimental psychology: A manual of laboratory practice* (Vol. 1). New York, NY: MacMillan.
- Torres-Manzanera, E. (in press). xkcd: An R package for plotting XKCD graphs. *Journal of Statistical Software*.
- Van Nes, F. L., & Bouman, M. A. (1967). Spatial modulation transfer in the human eye. *Journal of the Optical Society of America*, 57(3), 401–406.
- Vasari, G. (2007). *The lives of the most excellent painters, sculptors, and architects*. New York, NY: Modern Library.
- Viénot, F., Brettel, H., Ott, L., M'Barek, A. B., & Mollon, J. D. (1995). What do colour-blind people see? *Nature*, 376(6536), 127–128.
- Vogels, R. (1999). Effect of image scrambling on inferior temporal cortical responses. *NeuroReport*, 10(9), 1811–1816.
- von Ahn, L., Maurer, B., McMillen, C., Abraham, D., & Blum, M. (2008). reCAPTCHA: Human-based character recognition via web security measures. *Science*, 321(5895), 1465–1468.
- Wang, Z., & Simoncelli, E. P. (2008). Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities. *Journal of Vision*, 8(12):8, 1–13, doi:10.1167/8.12.8. [PubMed] [Article]
- Weegee & Speck, G. (1964). *Weegee's creative photography*. London, UK: Ward.
- Whitman, W. (1891). Eidolons. In W. Whitman & D. McKay (Eds.), *Leaves of grass* (pp. 12–13). Philadelphia, PA: David McKay.
- Wikipedia. (2016). *Duck typing*. Retrieved from [https://en.wikipedia.org/wiki/Duck\\_typing](https://en.wikipedia.org/wiki/Duck_typing)
- Young, R. A. (1987). The Gaussian derivative model for spatial vision: I. Retinal mechanisms. *Spatial Vision*, 2(4), 273–293.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353.

## Appendix A: Demo on publisher's website

We prepared a demo application (for PC; for Mac) that allows one to produce parameterized eidolons of various types for given images and save the results. The application also lets you conveniently look at the data structures that play decisive roles behind the screen. Thus, it shows more than is strictly necessary for eidolon generation. On the other hand, it is by no means a universal eidolon factory: There are things you might want to try that the demo doesn't allow. The reason is simply that we needed to keep the interface complexity—already complex—within reasonable bounds. We concentrated on instances that might prove of immediate interest to vision research.

Because the demo required an extensive interface, we economized on its capabilities. It lets you process only monochrome images that are 512 pixels square. Anything else will be converted to that format. The eidolon machine synthesizes the image on the basis of a scale-space representation of edgelets that are boundary representations that (different from the edginess obtained from edge detectors) retain the polarity along the boundary.

The demo will run on most platforms because it was written in Processing. For the applications, we packaged the java environment with the code so we can be certain that they will run properly regardless of what you may have installed on your machine.

The demo has an extensive interface that is convenient but perhaps takes getting used to. Most of the ins and outs are explained in the help that is always available under the “H” (or “h”) key.

## Appendix B: Pseudocode

### 1. Pseudocode

We find ourselves somewhat in a quandary about how we should describe the algorithmic aspects of the eidolon factory. Because proprietary aspects should not figure in a scientific journal, we cannot use a high-level formal language such as Mathematica (which would be most appropriate and clear from a formal perspective) or an environment such as Matlab (which might appeal more to engineering minds). Differences are substantial—for instance, one line of Mathematica may correspond to a thousand lines of C. Obviously, the former is easier to read than the latter. This is due to the hiding of details that are conceptually irrelevant.

Our implementation is in Processing, which is an open source Java-derived environment that was especially aimed at creative minds. Modern artists and designers use it extensively. We use it all the time in our vision research because it saves us so much time and effort, but we notice that few others in our field even know about it. Java runs on all platforms and is a well-designed object-oriented language. Processing might be “Java without tears” (for suckers), but it has retained these advantages. Almost any other language you might happen to be familiar with would serve just fine. Simply implement our pseudocode in your favorite language. This might (if you are at all familiar with your language) involve a few hours at most (as we actually checked!).

Unfortunately (or not; it depends on your perspective), there is no such a thing as a pseudocode standard. So we roll our own. It is perhaps something like formal pseudocode (i.e., not Fortran, Pascal, C, Basic, and so on style pseudocode). So you will see things like the following.

#### COMMENT

    Lorem ipsum dolor sit amet, ...

#### END COMMENT

DO increment counter **END DO**

Notice that ignoring what is in between the **COMMENT** and **END COMMENT** braces is not going to hurt you. It will have no consequences on the eventual algorithmic implementation. The idea is that the notation is self-documenting, so perhaps occasionally taking notice of comments might be a good idea, as they may have been inserted for some reason. But, in principle, **COMMENT** means that the lines until **END COMMENT** can be safely skipped. On the other hand, the **DO** part “increment counter” is crucial. Failing to increment the counter—whatever that may mean—is surely going to hurt you. A lead to what it might mean can often be gleaned from the context or comments. **DO** is followed by something that has to be done. Other comments come also as pairs of

braces, such as

```
FOR something running over such and so possibilities do something END
FOR,
or
```

```
IF some condition, do something ELSE do something else END IF
```

and so forth. We use indentation to highlight the inclusion structure.

A structure like

```
FUNCTION_NAME( parameters)
    compute something
RETURN result
```

is a function that encapsulates a computation. Notice that the parameter section might be empty, like for a function **COIN\_FLIP()** that returns `head` or `tail`. A program is a collection of functions. One of these is supposed to deliver the final result.

Of course, there are numerous decisions in implementations that certainly may make a difference but can hardly be counted to be our business. Here is a simple example. Images are represented in numerous ways in computer memories. But how the bit planes are ordered, and so forth, is not our business. A monochrome point (“pixel”) can be represented by a byte, integer, float, double, and so forth. This is not our business either, but it makes a difference. We will need Fourier-based methods. Whether one uses the latest fast Fourier transform (FFT; Heideman, Johnson, & Burrus, 1985) implementation, the old-fashioned Filon integration (Abramowitz & Stegun, 1972), or something rolled oneself is not our business. But, again, choices often make a difference! Often in computation time, sometimes in precision. They may have distinct limits of applicability. As said before, such technicalities will be skipped here.

## 2. The basic deterministic structures

The simplest representation is as a number of progressively blurred images at discrete scale levels separated by factors of two (coarse) or square root of two (almost always good enough). The highly blurred images may be subsampled spatially without significant loss, but this is usually inconvenient and memory is cheap.

For a  $512 \times 512$  image (for example), the range of scales would run from 1 (pixel size) to about 128 (one quarter of the image size). With a square root of two factor, that implies more than a dozen levels.

Building scale space implies

```
SCALESSPACE()
FOR scale is 1 to 128 by factors of square root of 2
    DO set scale-space level i to the blurred version of the image at
        that scale
    END DO
    COMMENT
        you may want to deal with boundary effects here. For
        instance, you may want to implement a circular, soft window.
        This probably involves starting with a larger image and
        sacrificing material near its boundaries. Although sometimes
        necessary, we don't think it'll be worth your while.
    END COMMENT
END FOR
RETURN the stack of scale-space levels
```

One typically uses FFT methods to do this (the demo uses JTransforms 2015), but Mathematica enables you to simply say “blur the image by so much,” which captures the conceptual content in a direct way. You may want to do some additional housekeeping here—for instance, handle boundary effects in some preferred way, subsample the highly blurred layers, and so forth. We put a few hints to such issues in the comments.



Building the difference scale space implies

```
DIFFERENCE_SCALE_SPACE()
  FOR all but the coarsest scale
    DO set the layer to the difference of the scale-space layer and
      the next coarsest layer
    END DO
  END FOR
  DO set the final layer to the final scale-space layer END DO
  comment
    - another method would be to simply compute Laplacians of the
      image and forget about the regular scale space.
    - Instead of saving the final layer, you may simply store the
      image mean intensity – it will hardly make a difference.
  END comment
RETURN the stack of layers
```

Notice that after constructing the difference scale space the scale space itself can be deleted because it can be regained from the difference scale layers.

One catch to be aware of is the DC level because DOG receptive fields are not sensitive to that. In practice, adding a constant suffices, so this is not a problem. You simply retain the coarsest scale-space layer. Thus:

```
SANITY_CHECK()
  DO add all layers of the difference scale-space END DO
  CHECK the result should equal the fiducial image END CHECK
  COMMENT
    Of course, only sissies would actually do this! Sanity checks will
    be deleted from the final implementation of your eidolon factory.
  END COMMENT
RETURN GO/NO_GO
```

### 3. Local disarray

The basic ingredient is the Gaussian noise image:

```
GAUSSIAN_NOISE_LAYER( parameter: scale )
DO generate an image with random pixel values END DO
COMMENT
    ideally you use a normal distribution with zero mean, unit
    variance. In practice, it makes little difference what you use,
    and speed may be more important. We use a good and fast PCG random
    generator (O'Neill in press).
END COMMENT
DO blur the image to a given scale END DO
COMMENT
    in order to speed things up you might generate a random noise
    SPECTRUM instead of a random noise IMAGE. Moreover, you may store
    blur kernel spectra instead of generating them fresh when you need
    them.
    Starting from a complex random spectrum yields two Gaussian fields
    for a single complex inverse FFT, it is probably the most
    efficient method.
END COMMENT
DO normalize the image to zero mean and unit variance END DO
COMMENT
    sometimes you need that, sometimes not. If you don't need it,
    don't do it. In generating stochastic signals you should generally
    normalize at the end only (prevents repeats and prevents bugs) and
    that only when necessary. Generating noise may well be a
    computational bottleneck, typically the normalization stage will
    dominate computation time.
END COMMENT
RETURN the image
```

The eidolon factory requires a great many of such images, all mutually independent. For instance, a displacement vector field requires two:

```
DISARRAY_FIELD( parameters: reach, scale )
DO generate two Gaussian random noise layers for the specified scale
END DO
DO package them as the Cartesian coordinates of a vector field END DO
DO scale these to the specified reach END DO
COMMENT
    This is one aspect that lets scale-space methods shine. You
    automatically generate isotropically distributed random vectors.
    The two-dimensional Gauss function is both isotropic and separable
    in Cartesian coordinates.
END COMMENT
RETURN the vector field
```

It is a simple matter to impose disarray on a given image:

```
DISARRAY_IMAGE( parameters: image, reach, scale)
  DO generate a disarray field of the specified reach and scale END DO
  DO allocate a new image END DO
  FOR each pixel of the new image
    DO set the pixel value to that of the pixel in the image at a
      location offset by the disarray field
  END FOR
  COMMENT
    - Notice that you can't do it the other way around (image to
      disarrayed image) because that would yield "holes", some pixels
      would not get a value.
    - Make sure that you don't look for pixels that are outside of the
      image. There are several methods to make sure this cannot happen.
      If the method makes a difference you're probably on the wrong
      track.
    - If you are the worrying type you may be inclined to do a bit of
      (say bilinear) interpolation here. We advise against that. It is
      slow and you won't be able to see the difference.
  END COMMENT
RETURN the new image
```

Notice that the “image” here will usually be a difference scale-space layer. You will perturb many such layers before combining them in the final synthesis. Notice also that there are many additional uses for noise fields. For instance, instead of or in addition to spatial disarray, you may want to perturb the gain, orientation, and so forth of a receptive field. Although we do not consider this in this article, here lies an important field of enquiry.

This is how you construct fractal disarray:

```
FRACTAL_DISARRAY FIELD()
  DO generate a stack of disarray fields for each scale value, setting
    the reach proportional to that scale value
  END DO
  FOR each stack layer, starting with the coarsest
    DO add the next coarse level if one exists END DO
  END FOR
  COMMENT
    Here you may want to add some kind of normalization to the
    results.
  END COMMENT
RETURN the stack
```



#### 4. Single-scale eidolons

It is easy enough to implement such eidolons. This is how (essentially just the regular CAPTCHA or XKCD-emulation method):

```
SINGLE_SCALE_DISARRAY( parameters: scale, reach, grain)
  DO assign a new image and set it to the blurred image at level scale
END DO
  DO generate a disarray field with the specified reach and grain END DO
  DO set the new image to the disarrayed copy END DO
COMMENT
  Since the image histogram remains invariant under disarray, there
  will be no need to normalize the resulting image.
END COMMENT
RETURN the disarrayed image
```

Indeed, nothing more complicated than that. The resulting image is both blurred and locally scrambled. Such eidolons are extremely simple to generate yet are already an interesting class for vision research—in fact, most likely a good starting platform in many cases.

### Appendix C: Relation to sparse coding and mongrels

The eidolon factory bears partial resemblance to the texture analysis/synthesis algorithm introduced by Portilla and Simoncelli (2000), which has been extensively used to produce “mongrel” pictures by Balas, Rosenholtz, and colleagues (Balas, 2006; Balas et al., 2009; Rosenholtz, 2011; Rosenholtz, Huang, & Ehinger, 2012; Rosenholtz, Huang, Raj et al., 2012).

The first stage of the Portilla and Simoncelli algorithm involves the decomposition of the original image into a series of oriented subbands plus an additional low-pass band—a so-called steerable pyramid (Simoncelli, Freeman, Adelson, & Heeger, 1992). From this decomposed representation, a set of descriptors are extracted, which are subsequently used to constrain the synthesis algorithm. The synthesis algorithm itself starts off with a random field of the same amplitude and variance as the source image and uses an iterative process to impose the constraints to the initial noise input. The constraints include marginal statistics of the image picture plus skewness and kurtosis as defined at a subband level, raw coefficient correlation, coefficient magnitude statistics, and cross-scale phase statistics.

This first striking difference between the eidolons and mongrels is that although the first are fundamentally characterized by two parameters, reach and coherence (which have identifiable perceptual correlates), the full description in the Portilla and Simoncelli algorithm is stored in a relatively large number of parameters (710 using the settings recommended in the original publication; Portilla & Simoncelli, 2000).

Although each constraint class has a functional meaning as a whole (e.g., the raw coefficient correlation parameters characterize the presence of periodic or globally oriented structures in the image), with the exception of the marginal statistics parameters, the meaning of the single predictors is not transparent. This means that it is hard to modify the descriptors in order to generate predictable perceptual effects unless one excludes one class of predictors altogether. It is, however, true that when considering constraints computed from natural images one can exploit the correlation patterns among predictors in order to make them more manageable. It has recently been shown that applying dimensionality reduction to the whole set of parameters can be used to reveal texture selectivity in V4 neurons (Okazawa, Tajima, & Komatsu, 2015).

The second important difference is that the eidolon factory works by perturbing the original image to create a new eidolon instance, whereas the Portilla and Simoncelli algorithm works by constraining a noise image to accommodate a given set of constraints. This means that the difference between a fiducial image and its eidolons can be reduced ad libitum by reducing the reach of the disarray, whereas the difference between the fiducial image and its mongrels is on average always the same and is determined by how well the descriptors capture the appearance of the specific texture. As a means to obtain mongrels that are nearer to the fiducial image, one can, however, seed the synthesis algorithm with the fiducial image corrupted by a certain amount of noise.

The third important difference between the eidolon factory and the texture synthesis algorithm is that the first does not use any iterative process. The Portilla and Simoncelli algorithm is to a large extent optimized for computing speed in particular as the pyramid repre-

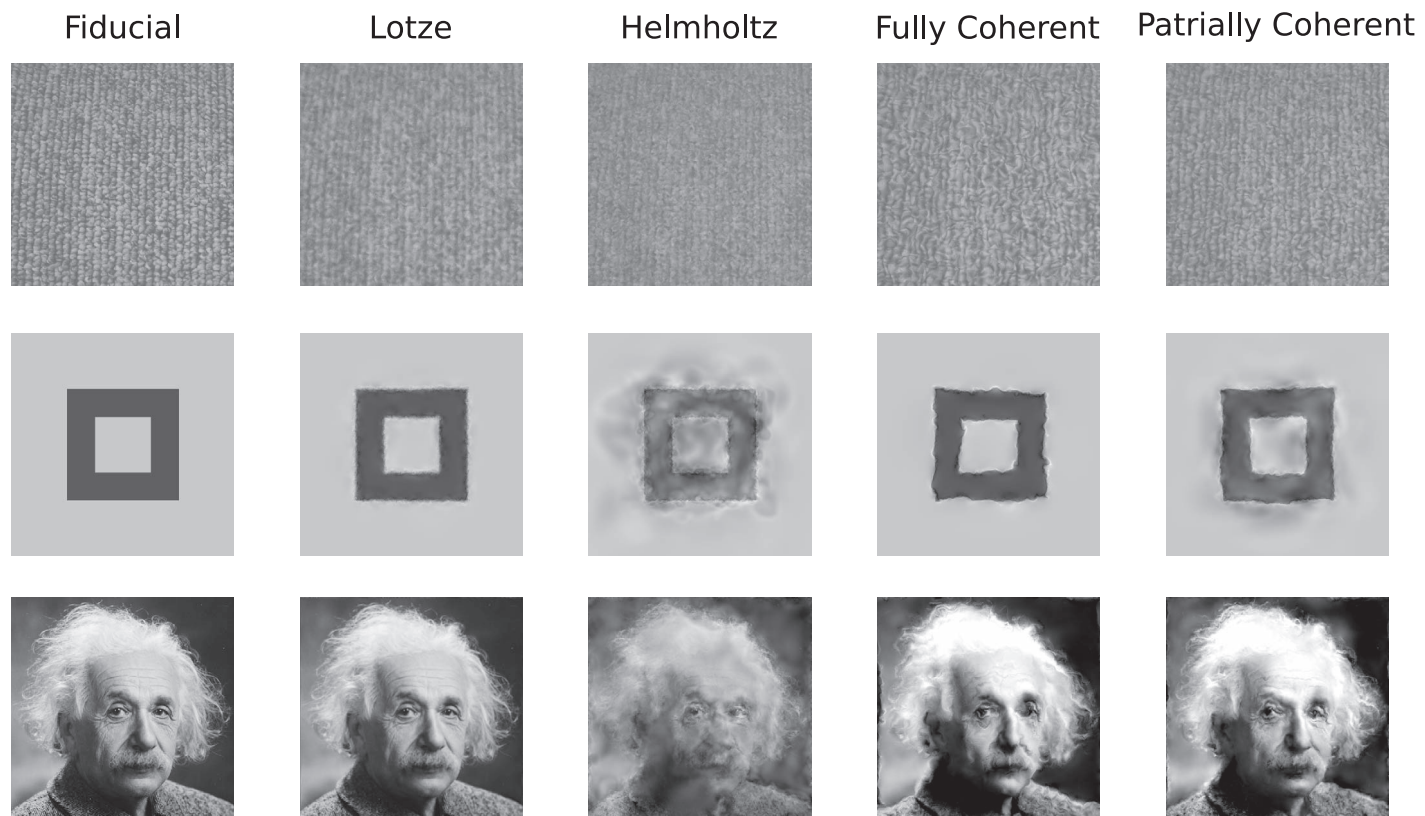


Figure 24. Examples of eidolons generated starting from a fabric texture image, a geometrical shape, and a face. Reach parameter was set at 0.5 for all examples. See the main text for the details about the different coherence configurations. Photograph of Albert Einstein reprinted from <http://www.loc.gov/pictures/item/2004671908/>. Copyright Orren Jack Turner.

sensation of scale space is very economical (Burt, 1981). The eidolon factory is bound to be faster, if anything because the same disarray field can in principle be applied to any number of fiducial images as long as they have the same size. On the contrary, the texture optimization procedure has to be repeated for each fiducial image and random seed.

Figures 24 and 25 show examples of eidolons and mongrels obtained with different parameter settings and different types of seeds, respectively. A quick glance at the different examples easily shows the strengths and weaknesses of the two approaches when it comes to generating images that are perceptually equivalent to the fiducial image. The eidolon factory, given a reasonable reach value, performs much better at preserving the topology of the fiducial image, which is very evident when comparing the eidolon generated from the square image with full coherence and the mongrel generated with a fully random seed. At the same time, the Portilla and Simoncelli algorithm performs very well at preserving the periodic fine-scale patterns in the fabric image and the general vertical–horizontal orientation of the sharp edges in the square image, which are to a large extent altered by the eidolon generation.

Certainly, an appropriate configuration of the parameters in the two algorithms coupled with the appropriate fiducial image can produce results that look relatively similar in a nontrivial way—that is, when the synthetic images are different from the fiducial image but the overall image structure is preserved (see Figure 26 for an example). We suspect, however, that to a trained observer familiar with the fiducial image the two classes of synthetic images would still be distinguishable because the Portilla and Simoncelli algorithm introduces perturbations at random locations, as new edges and gradients appear solely due to the initial random field. The features in the eidolons instead remain identifiable in the vicinity of their original location.

## Appendix D: Glossary

**Differential geometry:** Local geometry. It is defined by being applied to regions of interest that have the same size as the operators (e.g., edge detectors).

**Edgelet:** Local component of an edge. In differential geometry, edges are considered as a string of spatially contiguous and aligned edgelets.

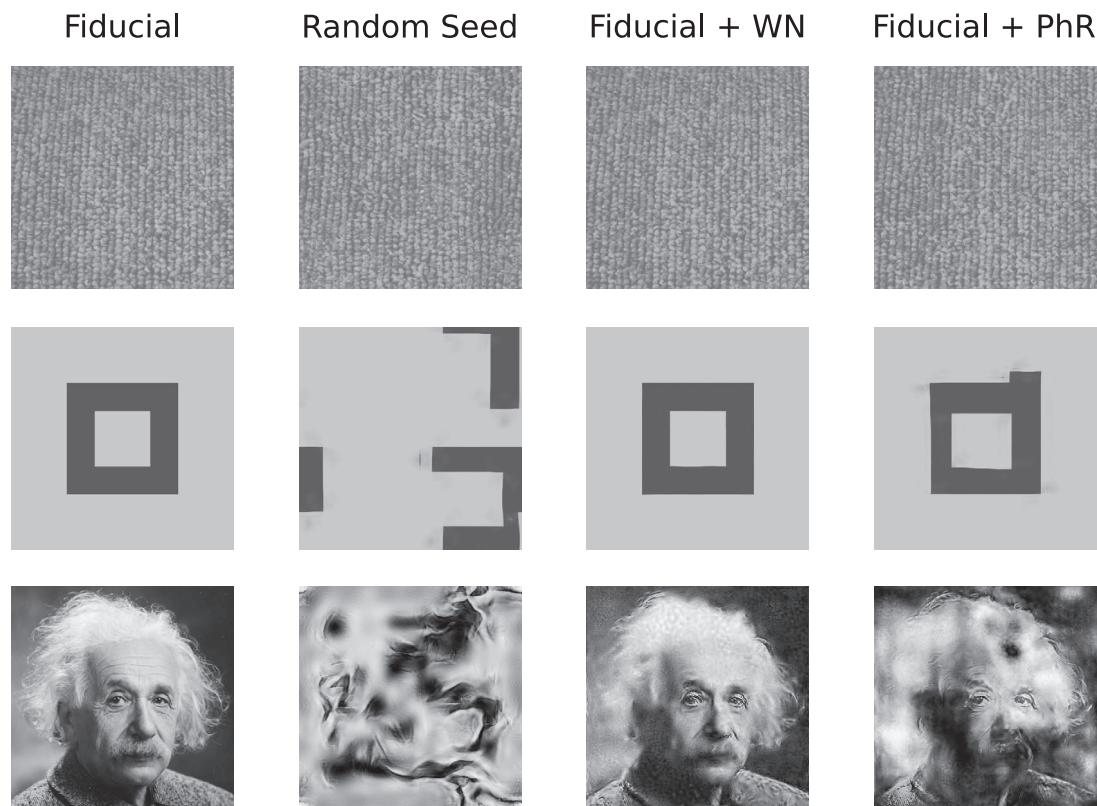
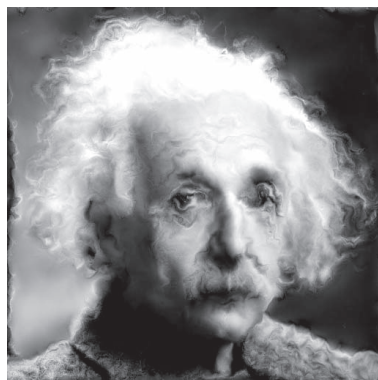


Figure 25. Mongrels of the same three images generated with the Portilla and Simoncelli algorithm. The images have been generated using either a phase-randomized version of the original image as a seed or an equal mixture of the original image and its pixel-scrambled version (WhN) or its phase-scrambled version (PhR). Photograph of Albert Einstein reprinted from <http://www.loc.gov/pictures/item/2004671908/>. Copyright Orren Jack Turner.

**Eidolon:** Class of stimuli that are equivalent to a given fiducial stimulus along a given perceptual continuum. Stimuli that are metameric are eidolons too, but the definition extends to stimuli that are

perceptually equivalent along a given dimension while still being distinguishable in other aspects. Notice that equivalence is defined in a phenomenological sense, and consequently it is subjective in nature.

## Eidolon



## Mongrel

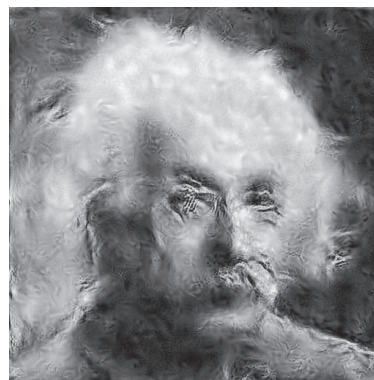


Figure 26. Comparison of an eidolon (obtained with fully coherent disarray and 0.5 reach) and a mongrel (seed corrupted with high-passed noise). Even if the two instances look relatively similar, some qualitative differences are evident, such as the fact that features (e.g., edges) can appear or be enhanced at random positions in the mongrel image, whereas each feature can be traced to the fiducial image in the eidolon. Photograph of Albert Einstein reprinted from <http://www.loc.gov/pictures/item/2004671908/>. Copyright Orren Jack Turner.



**Eidolon factory:** Algorithm that can be used to modify images. Its parameterization defines the physical space in which perceptual equivalence can be established through psychophysical methods. Many eidolon factories are possible beyond the one we introduce in this work.

**Local sign:** Positional signature (German *Lokalzeichen*). Psychophysical bridge between neural representation and awareness of position.

**Metamer:** Class of stimuli that are perceptually indistinguishable under some specific viewing condition.

**Modulation transfer function (MTF):** Being  $\nu$  a given spatial frequency of a grating stimulus,  $C_0$  the physical contrast of a stimulus, and  $C_i$  the transferred contrast (i.e., the contrast resulting after the stimulus is

transferred through an optical device or the effective contrast in a visual system),  $MTF(\nu) = C_i/C_0$ .

**Psychogenesis:** Process by means of which a mental state comes to be. In the present study, we are primarily referring to the process leading to visual awareness when human observers view an image.

**Tarachopia:** Scrambled vision. Concept proposed by Hess (1982) to characterize the phenomenology of amblyopia as well as the observation that amblyopia affects pattern discrimination to a larger extent than simple visual detection.

**Translation invariance:** Indicates that the measurement of a property is independent of the location at which the measurement takes place. Specifically, in the case of Fourier transform, it indicates that the amplitude spectrum is identical if the image is shifted.