

The Independent Components of Natural Images are Perceptually Dependent

Matthias Bethge, Thomas V. Wiecki and Felix A. Wichmann

Max Planck Institute for Biological Cybernetics, Dept. of Empirical Inference,
Spemannstr. 38, 72076 Tübingen, Germany.

ABSTRACT

The independent components of natural images are a set of linear filters which are optimized for statistical independence. With such a set of filters images can be represented without loss of information. Intriguingly, the filter shapes are localized, oriented, and bandpass, resembling important properties of V1 simple cell receptive fields. Here we address the question of whether the independent components of natural images are also perceptually less dependent than other image components. We compared the pixel basis, the ICA basis and the discrete cosine basis by asking subjects to interactively predict missing pixels (for the pixel basis) or to predict the coefficients of ICA and DCT basis functions in patches of natural images. Like Kersten (1987)¹ we find the pixel basis to be perceptually highly redundant but perhaps surprisingly, the ICA basis showed significantly higher perceptual dependencies than the DCT basis. This shows a dissociation between statistical and perceptual dependence measures.

Keywords: redundancy reduction, independent component analysis, information theory, natural image statistics, psychophysics

1. INTRODUCTION

Visual perception starts with two-dimensional (2-D) arrays of light falling on retinae—at least for mammals. The task of visual perception is to enable the animal to use the information provided in the array of light in order to react appropriately to the (3-D) objects surrounding it. Clearly, this is a problem of *inference*, as already pointed out by Helmholtz in the middle of the 19th century. The problem of understanding vision—that is, to understand how animals and humans not only solve this “inference” but usually do so accurately, quickly and seemingly effortlessly—is a formidable task and has attracted generations of thinkers from philosophy, mathematics, physics, biology, psychology, and, more recently, computer science. For spatial vision and object perception, understanding the “inference” should allow us to answer Koffka’s² deceptively simply sounding central question “Why do things look as they do?” (Koffka, 1935, p. 76).

1.1. Early Vision and Linear Systems Theory

The starting point of visual perception, and the starting point to understand the inference process, is the initial encoding of light on the retina and the translation of the retinal image into neural representations of the visual world. This field of research is often referred to as *early vision* or *low-level vision*. Mimicking the experimental approach so very successful in physics, the study of visual processes has traditionally been directed at isolated physical dimensions of visual information such as luminance, colour, or motion. In this paper we are only concerned with the initial encoding of luminance, that is, we are concerned with early *spatial* vision. Many classic, important, and frequently exceptionally ingenious and elegant experiments have been carried out in this domain.³⁻⁹

Campbell and Robson’s (1968)⁷ seminal paper argued that—at least some—visual stimuli may be best represented in the Fourier domain in order to be able to predict their detectability or discriminability for human

Further author information: (Send correspondence regarding image statistics and computations to M.B., and correspondence regarding psychophysics to F.A.W.)

M.B.: E-mail: mbethge@tuebingen.mpg.de, Telephone: +49 7071 601-1770

F.A.W.: E-mail: felix@tuebingen.mpg.de, Telephone: +49 7071 601-554

observers. This perhaps counter-intuitive claim implies that the representational format for visual stimuli may not always resemble phenomenology. Furthermore, their results suggested that the early visual system's essential properties could be captured in a linear system—the now ubiquitous linear, band-limited “spatial-frequency channels.” Ever since Campbell and Robson's paper a multitude of psychophysical experiments explored the response of the visual system to simple stimuli with carefully chosen properties in the Fourier domain hoping to find the (linear) transfer function equivalent of the early visual system. However beautiful some of the experiments and their results may be, they almost rather obscured the central question: What is vision for? Can we find general basic principles of the initial inference process? Can most of the research on the early stages of information processing in the brain be subsumed under a more general heading, so as to help us connect the seemingly isolated pieces of knowledge into a coherent picture?

As Campbell and Robson argued, in spatial vision it is often fruitful to consider the stimuli in the Fourier domain as the visual system appears to initially code the world via bandpass filters. Motion detection, on the other hand, is likely accomplished by local delay-and-compare operators, the disparity between images falling on the retinae can be used to recover depth information, and colour is coded in colour-opponent channels. All of these operations can be concisely and elegantly subsumed under a common heading: taking measurements via blurred derivatives (see the exemplary exposition by Adelson and Bergen¹⁰). The first and most basic task of vision is to measure—estimate—what is out in the world. Mathematically, function estimation is well accomplished by taking local derivatives; however, whilst in mathematics the derivative is taken at a point only, for real-world applications with noise—and if the functions to be estimated are smoothly varying—we gain more robust information if we take a local average (blur) and then the derivative. Since both blurring and taking the derivative are linear operators, we can fuse both operations into a single blurred derivative operator. Blurred derivative operators resemble simple cell receptive fields reasonably well. Of course, different formalisations of simple cell receptive fields as Gabor functions, differences of Gaussians (DoG's), or derivatives of Gaussians, and many others exist. However, all of them can be viewed as blurred derivative operators as long as they contain inhibitory and excitatory subregions.

1.2. Early Vision and Natural Image Statistics

How complex or simple a structure is depends critically upon the way we describe it. Most of the complex structures found in the world are enormously redundant, and we can use this redundancy to simplify their description. But to use it, to achieve the simplification, we must find the right representation.

Herbert A. Simon^{11,12}

In the last section we argued that it is beneficial to view the first task of any sensory system to measure what is in the world via (blurred) derivatives, that is, we took a *functional* stance to look at early vision. However, one may also adopt a *normative* stance: how a sensory system *ought* to be organized from an engineering or information-theoretic point of view.

The relationship between the raw measurements of light intensities in the retina and the behaviorally relevant variables to be extracted is very subtle. In particular, the response of any photoreceptor viewed in isolation from the others cannot tell anything about the spatial structure of a scene or the presence or absence of objects. Scenes and objects—the content of an image—can only indicate themselves through characteristic *patterns* hidden in the high-dimensional signal provided by the multitude of receptor responses.

Attneave¹³ and Barlow¹⁴ were among the first to hypothesize that there should be a connection between the statistical regularities of the visual world and our visual system: Our visual system—the neural representations—should be adapted to the statistics of “natural images.” With sufficiently powerful desktop computers now readily available, the study of the general statistics of natural images has become a topic of growing interest. Typically, researchers attempt to relate the statistical properties of natural scenes to the processing of the early stages of the visual system^{1,15-21}—for a review see Simoncelli and Olshausen.²² The hypothesized connection between the statistics of natural images and the properties of sensory neurons in the visual system comes from information theory: efficient coding and redundancy reduction.

Natural images are highly redundant, that is, they exhibit large correlations between neighbouring pixels because the world consists of (partially occluded) objects, and nearby regions of the same object or surface are

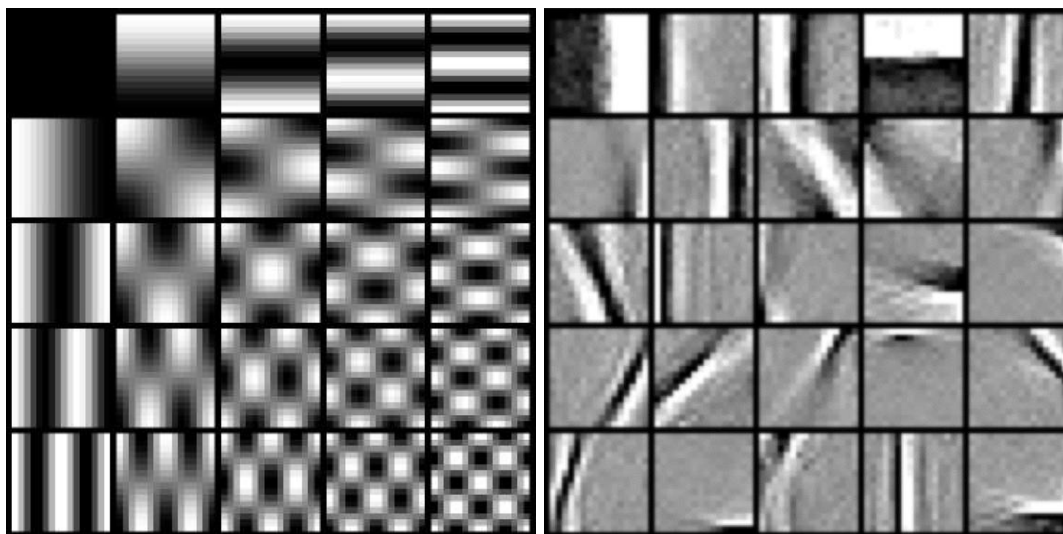


Figure 1. The DCT basis functions (left) and the ICA basis functions (right) used in the psychophysical experiment. The ICA basis has been optimized for the van Hateren images²³ (see Fig. 2).

very likely to share the same luminance (and colour). Coding natural images point-by-point, or pixel-by-pixel, is thus a very inefficient representational format. At the level of the photoreceptors this is, however, precisely how the visual system represents its input. The redundancy reduction hypothesis suggests that the fundamental computations of the early stages of the visual system are influenced by the goal of statistical independence.

1.3. Second-Order vs Higher-Order Redundancies

Second-order correlations between pixel intensities are defined by the expectation $E[I(x, y)I(x', y')]$. Using principal component analysis (PCA), it is always possible to find an orthogonal linear transformation whose output coefficients are completely decorrelated. At least for local image patches, it is plausible to assume a translation invariant statistics for the pixel intensities, meaning that $E[I(x, y)I(x', y')] = C(x - x', y - y')$. In this case, the principal components are not unique but always resemble products of one-dimensional sine functions in the x and y direction. Well known examples include the Fourier basis and the basis of the *Discrete Cosine Transform (DCT)*, which both play a fundamental role in image processing. For redundancy reduction, the Discrete Cosine Transform is of particular interest as it has proven useful for image compression in the widely used JPEG standard (see Fig. 1 left).

The assumption of second-order decorrelation alone is not strong enough for deciding which representation to use for images. In addition to the modest ambiguity in choosing the principal components mentioned in the previous paragraph, this ambiguity becomes excessive when the restriction of PCA to orthogonal transforms is dropped.^{24,25} Therefore, it is a natural extension from the perspective of the redundancy reduction hypothesis to optimize the representation for higher-order correlations as well.

In an influential paper, Olshausen and Field (1996)²⁰ were able to demonstrate the use of higher-order statistics for learning an image basis: By training a neural network to sparsely code natural image patches they found that the resulting basis functions were localized, oriented, and bandpass—features also shared by V1 simple cell receptive fields. Later, several researchers applied variants of *independent component analysis (ICA)* to natural images.²⁶⁻²⁹ Applied to natural images ICA, too, results in basis functions very much like those of V1 simple cells (see Fig. 1 right).

Like DCT, ICA is based on a linear transformation which makes both representations directly comparable: The only difference between the two transformations lies in the choice of a different basis (for comparison see Fig. 1, left vs right).

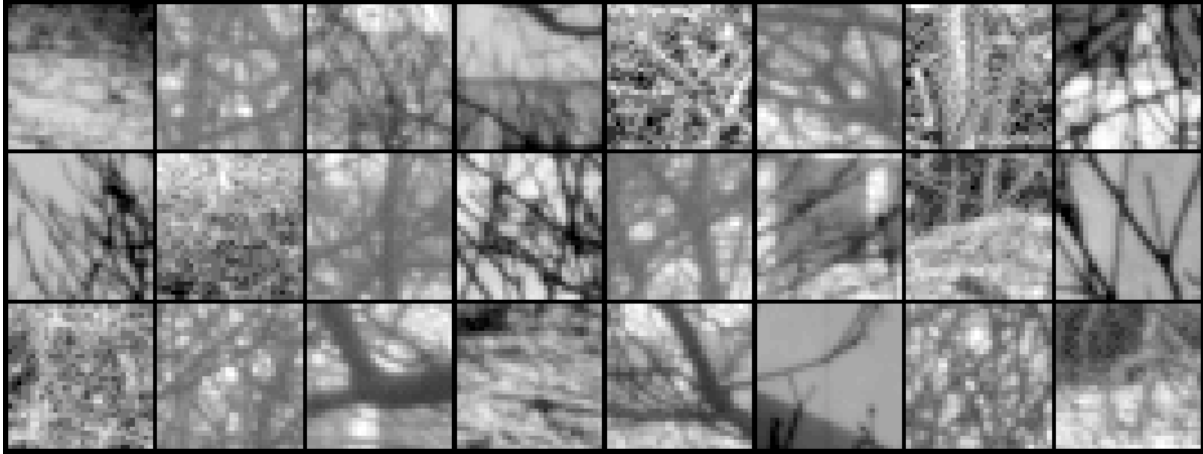


Figure 2. Natural image patches used. The patches were extracted from images of the van Hateren natural still image collection.²³

Given that we are still bound to the linear model, how important is the optimization for higher-order dependencies? Linearity strongly regularizes the class of possible mappings. Therefore, it is not obvious whether the reduction of higher-order redundancies can actually make a large difference in a linear model. In fact, it has been shown by Bethge²⁵ that the gain of the ICA basis over any other decorrelation basis in terms of redundancy reduction is only of the order of five percent or even smaller.

Regardless of the small information-theoretical improvement of the ICA basis over the DCT basis (or any other decorrelating representation), this difference may nonetheless be important empirically. Thus, we complement the theoretical investigations of Bethge²⁵ with a direct experimental approach in order to test the perceptual significance of different bases.

2. PSYCHOPHYSICAL EXPERIMENT

Direct psychophysical measurements of the redundancies in image representations are rare, with the notable exception of Kersten's study¹ which shows how *perceptually*—and not only statistically—redundant the pixel representation of images is. Kersten's study utilized Shannon's guessing game³⁰ to estimate the entropy of images. Similarly, our study also instantiates a guessing game, yet we take a less ambitious paradigm which does not aim at estimating the correct *amount* of entropy. Rather we would like to perform a robust test against the null hypothesis of perceptual independence, which considerably simplifies the nature of our guessing game: the better subjects can predict, the less likely is the null hypothesis, the more significant is the redundancy.

2.1. Experimental Setup

The stimuli—24 patches of 32 x 32 pixels were randomly sampled from ten images of the van Hateren natural still image collection²³.

Stimuli were presented against the mean luminance (213 cd/m²) of a carefully linearised Siemens SMM21106LS gray-scale monitor driven by a Cambridge Research Systems Visage display controller (spatial resolution 1024 x 768 pixels at a refresh rate of 130 Hz non-interlaced); they were presented within a rectangular temporal envelope for however long an observer wanted (see section 2.2 for details). At the viewing distance of 50 cm the whole display nominally subtended 43 degrees of visual angle; individual patches in the 2x3 array of image patches subtended 6 degrees (again see section 2.2 for details).

Eight observers with normal or corrected-to-normal vision acted as experimental subjects; two of them were authors (MB and FAW), the others were naïve to the purpose of the experiment and were paid for their participation. Each observer played the guessing-game, described below, for 24 patches for pixel-, ICA- and

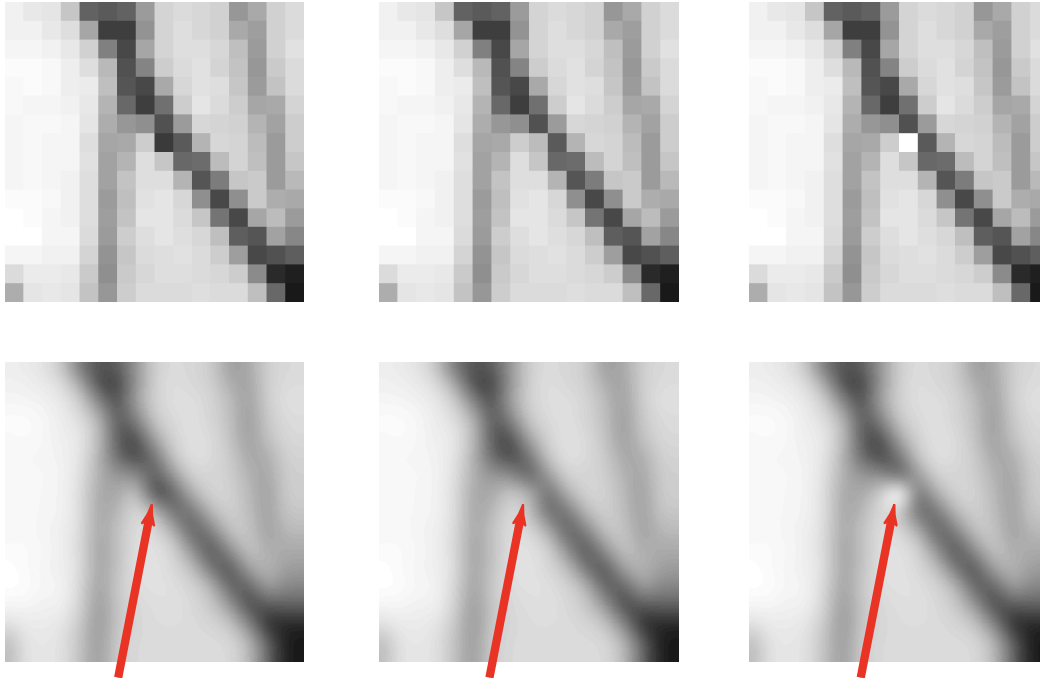


Figure 3. Example stimulus for the pixel-guessing-game; the three columns correspond to the maximally different pixel intensities for the pixel basis; observers always saw patches in their native resolution of 16x16 pixels (top row) and low-pass filtered to remove the blocking artifacts (bottom row). The target pixel is indicated by the red arrow in each bottom-row panel.

DCT-bases, for a total of $(4 + 1 + 1) \times 24 = 144$ judgements per observer (the pixel adjustment task was carried out for 4 different patch sizes).

2.2. Psychophysical Method: Nested Interval Adjustment Scheme

The nature of the guessing-game necessitated the development of a new psychophysical method to obtain reliable data within a reasonable amount of experimental time. In pilot experiments we found, e.g., that the method of adjustment worked well for the pixel-basis but led to highly variable and painfully slow trials for the ICA- and DCT-basis.

The data reported here came from a *Nested Interval Adjustment Scheme* where subjects were always presented with three different versions of the same image patch. In the following we first explain the Nested Interval Adjustment Scheme for the pixel guessing-game, as this is the most intuitive. In the original Kersten-style pixel guessing-game,¹ the pixel intensities have been resolved with 4 bit resolution which gives rise to 16 different gray level-levels. Observers saw an image with one pixel set to a randomly determined gray-level, and the observers were asked to set the pixel to its “true” value (a discrete variant of the method of adjustment).

In our experimental paradigm it is not necessary to choose a fixed resolution for the gray levels. The images were always displayed with maximal resolution of 12 bits which corresponds to 4096 different gray levels. Observers were presented with three versions of an image patch (see Fig. 3): Instead of adjusting the gray-level of the experimental-pixel they selected the one closest to the “true” gray-level. If we encode for the whole range of possible gray-level values by rational numbers between 0 (‘black’) and 1 (‘white’), the first three alternatives in the pixel guessing-game would always be $1/2 - 1/3 = 1/6$ (‘dark’), $1/2$ (‘mid-gray’), and $1/2 + 1/3 = 5/6$ (‘bright’), respectively. Assume they had selected the mid-gray pixel. The next iteration (or level) of this trial then presented them with pixel values $1/2 - 1/6 = 1/3$ (‘light-gray’), $1/2$ (‘mid-gray’) or $1/2 + 1/6 = 2/3$ (‘dark-gray’) in the three image patches. Again the observer selected the one they felt was closest to the true

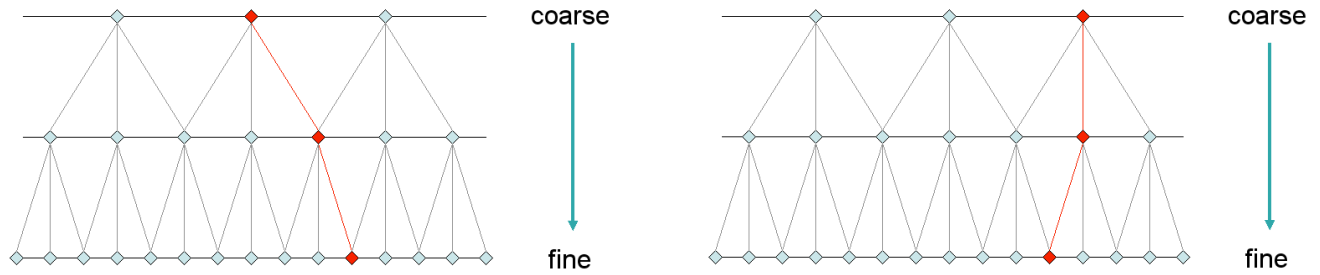


Figure 4. Graphical demonstration of our nested interval adjustment scheme. The two examples (left vs right) show that, unlike in a simple bisection setting, different trajectories through the stimulus space can lead to the same outcome—the same perceptual judgement by our subjects. Subjects stopped a trial when any of the three options at a lower level—smaller differences—of the tree-like structure looked indistinguishable to them.

pixel value. Observers continued to a lower level—smaller and smaller differences—until the three patches looked indistinguishable to them. (Typically observers required 7.3 ± 2.1 iterations for the pixel-, 6.7 ± 1.7 for the ICA- and 7.1 ± 2.4 for the DCT-basis.)

Figure 4 shows a stylised version of the pixel guessing-game. After choosing one—the mid-gray one in Figure 4a, the black one in Figure 4b—the observer is presented with three further choices. In Figure 4a the observer chooses the darkest alternative, in Figure 4b the middle alternative. In the third iteration the observer in Figure 4a again chooses the darkest alternative, whereas the observer in Figure 4b now chooses the lightest alternative. Importantly, they both end up at exactly the same estimate of the true pixel intensity despite their different choices at both prior (“upper”) levels of the Nested Interval Adjustment Scheme. Thus this scheme allowed observers to recover from sub-optimal first choices, unlike a strict interval bisection scheme.

In principle, the guessing-game for the ICA- and DCT-basis functions worked in exactly the same way, only that the *coefficients* of the basis functions could be adjusted: the three options for a given basis function of a given basis set were 1.0 (maximal), 0.5 (mid-level) and 0.0 (minimal) at the first level, and so on. Figure 5 shows an actual example from the experiment for the ICA basis. Two aspects of Figures 3 and 5 are noteworthy: first, we actually presented each patch six and not three times; once at its native resolution of 16×16 pixels (top row) and once low-pass filtered to remove the blocking artifacts (bottom row). Observers were instructed to inspect whichever version they preferred to arrive at their judgement. The arrows mark the area of the patches where the change of this particular ICA basis coefficient results in the largest (perceptual) change in the image patch.

An important difference between the guessing-game with pixels and the guessing-game with the ICA or DCT basis is the shape of the basis functions: For the pixel basis, the different basis functions differ only by the locations of the pixel but not by their spatial shape. In contrast, for both, the ICA as well as the DCT basis, the different basis functions exhibit important differences in the shape. Since it is not feasible to test every basis function for a given image patch, it raises the question which basis function should be made experimental? We decided to always choose the basis function whose coefficient is largest in amplitude. This seems to be a well-suited choice for two reasons: First, a large coefficient indicates that the basis function contributes to the perceived structure in a meaningful way. Thus, it is more likely that the selected image component is not perceptually discarded as noise. Secondly, the selection bias towards large amplitude coefficients gives rise to a bimodal *a priori* distribution of the coefficients with vanishing density at and around zero. The advantage of such a distribution is that it penalizes conservative guessing strategies because here a guessing bias towards zero does not pay off.

For the pixel basis, the guessing-game was carried out for four different patch sizes: first only a small part (4×4 pixels) at the center of the patch was presented. Subsequently the aperture was increased to 8×8 , 16×16 , and finally to full patch size of 32×32 pixels. Thus the amount of context information that could be used by the observer for the pixel prediction task was successively increased.

For the ICA and DCT basis, the image patches were presented only at the size of 16×16 pixels. On average,

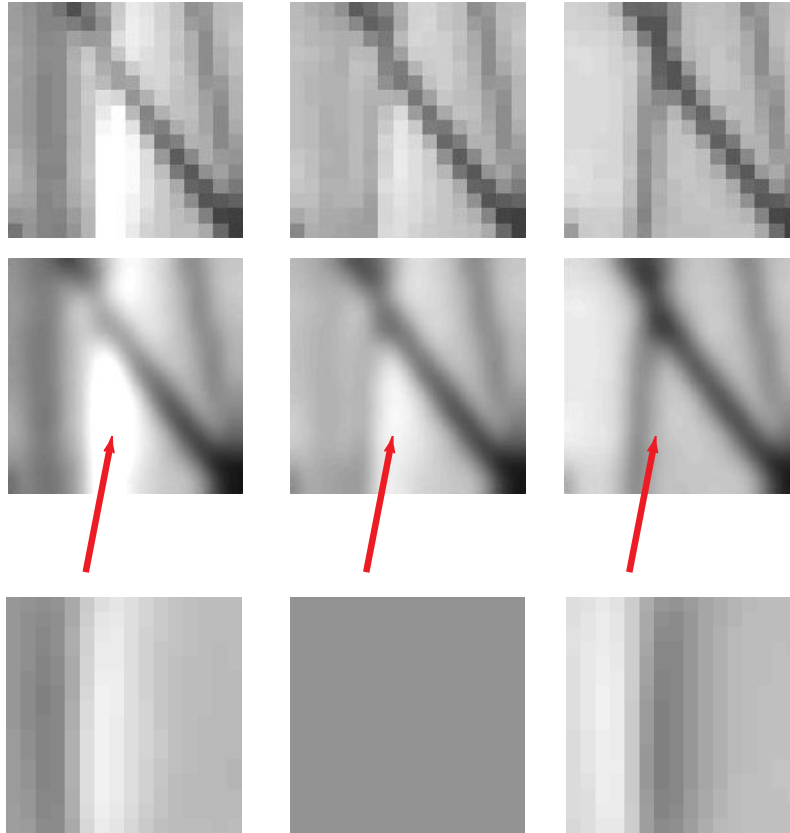


Figure 5. Example stimulus with the same image patch as in Fig. 3, however, now the three columns correspond to three different coefficients of an ICA basis function. The lower panel shows the ICA basis function multiplied by the three different coefficients as used in the image patches above.

observers required 7.3 ± 2.1 seconds for a pixel-basis judgement; the times were 30.4 ± 24.1 seconds for the ICA- and 23.9 ± 14.8 seconds for the DCT-basis.

3. RESULTS

The presentation of the results is divided into two classes: the pixel prediction task with four different patch sizes and the ICA/DCT coefficient estimation task for a fixed patch size. For each condition, we show the raw data by plotting the coefficients estimated by the subjects over the ground truth. In addition, we compute the correlation coefficients and the p-values for the null hypothesis that the data are uncorrelated. Finally, we also analyze the correlation between the signs of the estimated and the true coefficient.

3.1. Guessing-game with Pixels

Only four of the eight subjects participated in the guessing game with pixels (BH, FW, AF, MB). The raw data are shown in Fig. 6. The correlation coefficients shown in Fig. 8 (left panel) indicate that our observers were very good indeed at predicting the pixel values ($r^2 \geq 0.8$ for all observers and conditions). In fact the p-values—that observers were simply guessing—are extremely small ($p_{4 \times 4} = 8.3 \cdot 10^{-59}$, $p_{8 \times 8} = 3.7 \cdot 10^{-81}$, $p_{16 \times 16} = 1.6 \cdot 10^{-75}$, $p_{32 \times 32} = 4.5 \cdot 10^{-72}$ computed over all four subjects). Remarkably, the average correlation between guess and ground truth increases only between the patch sizes 4×4 and 8×8 (although the pixel size has been kept constant). This shows that the observers make use of very local information only.

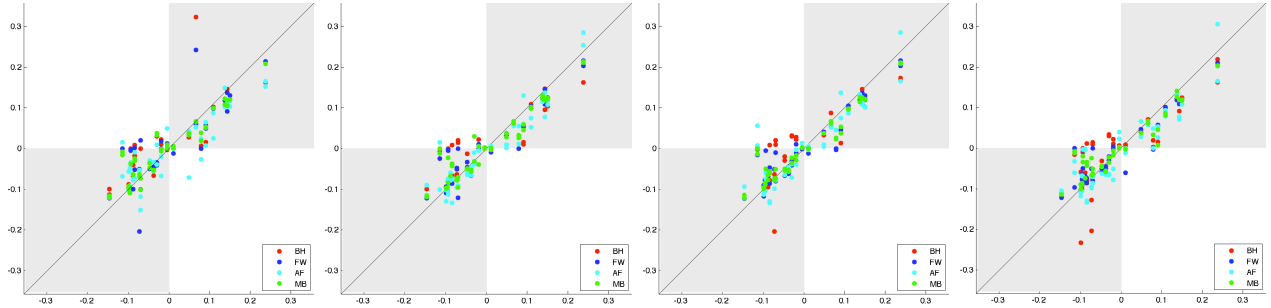


Figure 6. Results for the guessing-game with pixels. Each plot shows the subject’s guess (y-axis) as a function of the true pixel value (x-axis) using linear coordinates. The four different plots from left to right correspond to the four different patch sizes used (from left to right: 4×4 , 8×8 , 16×16 , and 32×32 pixels). The shaded quadrants indicate for which data points the sign has been predicted correctly.

3.2. Guessing-game with ICA and DCT

The central result of this paper is the finding of highly significant correlations between the coefficients of the ICA basis. While the estimate of the magnitude is very noisy in both cases (Figure 7, top row), subjects are significantly better in predicting the signs of the coefficients in the case of ICA compared to the case of DCT (Figure 7, bottom row): First, we found that the probability of ‘passes’ (i.e. the subjects set the coefficient to zero) is twice as large in the case of DCT ($P_{DCT}(\text{‘pass’})=13 \pm 2\%$) than in the case of ICA ($P_{ICA}(\text{‘pass’})=6 \pm 2\%$). Furthermore, given a nonzero coefficient has been selected, the probability of guessing its sign correctly was significantly larger in the case of ICA ($P_{ICA}(\text{‘correct’}|\text{‘no pass’})=78 \pm 3\%$) compared to the case of DCT ($P_{DCT}(\text{‘correct’}|\text{‘no pass’})=64 \pm 4\%$).

The prediction performance of the signs can be used to compute an information-theoretical lower bound on the redundancy by computing the information rate between subject guess and ground truth per coefficient. This information rate can only underestimate the true redundancy between the pixel intensities because the perceptual “take-up” of the pixel intensity information is limited. For the DCT basis, we find a perceptual redundancy of 0.048 ± 0.018 bits/coefficient, while in the case of the ICA basis we obtain a six times larger rate of 0.234 ± 0.039 bits/coefficient (see Figure 7, bottom row). Since the maximally possible information rate for the sign information is 1 bit/coefficient, we have a *redundancy of at least 23% for the ICA basis*.

The differences in the performance in predicting the signs of the basis function coefficients are also clearly visible from the correlation coefficients (Fig. 8, right panel). In particular the p-values exhibit a striking difference: The probability of observing the data under the null hypothesis of zero correlation between the experimental coefficients and the rest is $p_{DCT} = 0.2$ for the DCT basis. Hence, there are no statistically significant correlations in this case. In contrast, the p-value for the ICA basis is as small as $p_{ICA} = 1.2 \cdot 10^{-11}$, which strongly suggests the presence of statistically significant correlations according to standard hypothesis testing procedure.

4. DISCUSSION AND CONCLUSIONS

Our central finding is that the ICA basis exhibits significant, perceptually measurable redundancies while this is not the case for the DCT basis. This finding may be surprising because by definition the ICA coefficients are the least redundant coefficients possible that can be achieved using a *linear* transform. In order to understand this result we have to understand both: first, the limitations of ICA with respect to redundancy reduction due to the constraint of *linearity* as well as, second, the difference between redundancies among pixel intensities in an image and *perceptual redundancies* due to the loss of information during the process of perception.

Note that the difference between ICA and second-order decorrelation lies only in the objective function: While PCA minimizes the log-variances of the different components, ICA seeks to minimize the (marginal) entropies of the coefficients. Both objective functions are equivalent in case of Gaussian distributions while in the case of ICA also the deviation from a Gaussian in the shape of the coefficient histograms matters. Although the objective

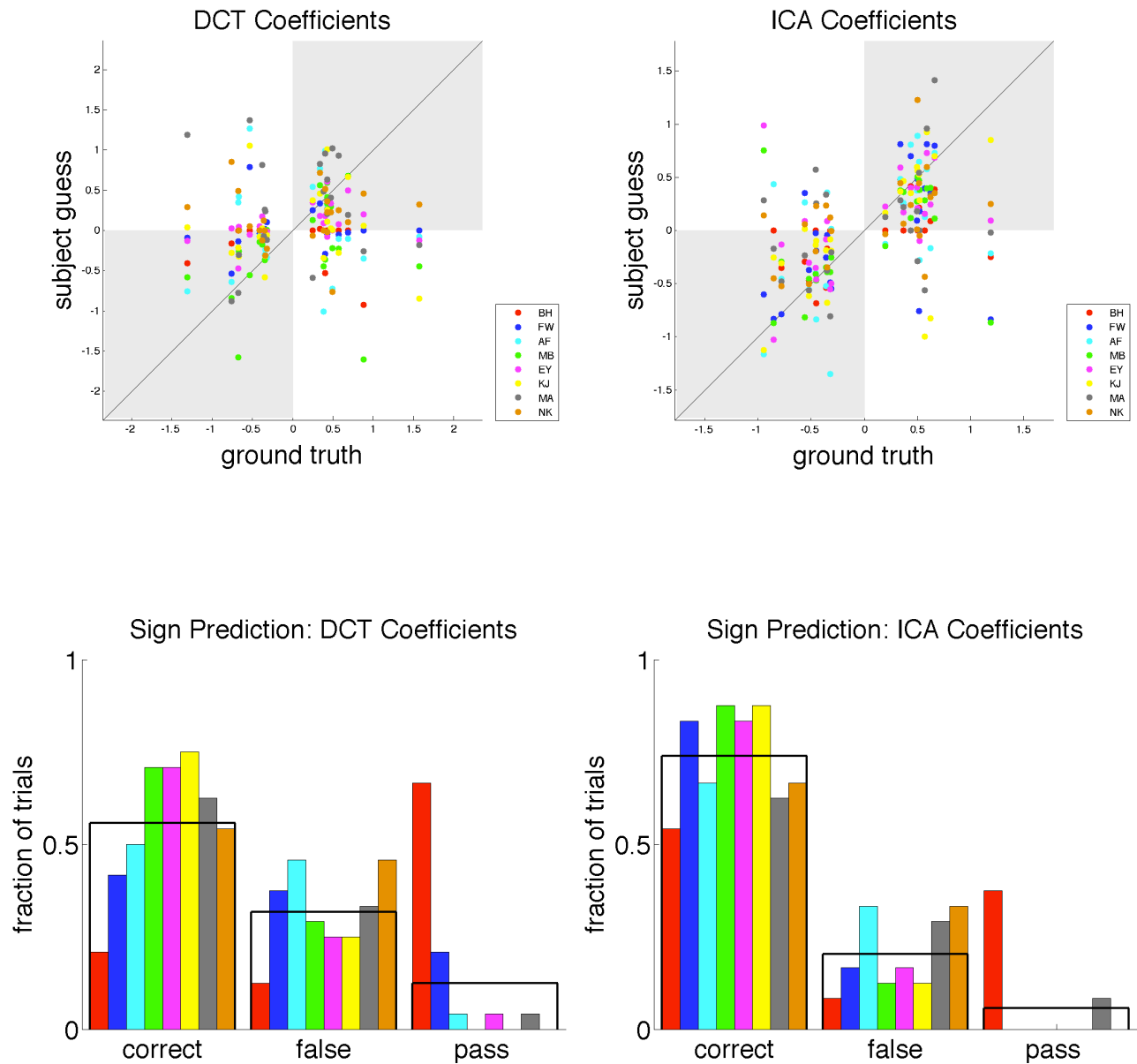


Figure 7. Results for the guessing-game with the DCT basis (left) and the ICA basis (right). Upper panel: Both plots show the subject’s guess (y-axis) as a function of the true pixel value (x-axis) using linear coordinates. The shaded quadrants indicate for which data points the sign has been predicted correctly. Lower panel: Bar charts indicate the fraction of trials for which the sign of the coefficient has been predicted correctly. If the estimated coefficient is zero, it has been classified as ‘pass’. Data for individual observers shown as narrow coloured bars, mean across observers as wide, transparent bars with black contours.

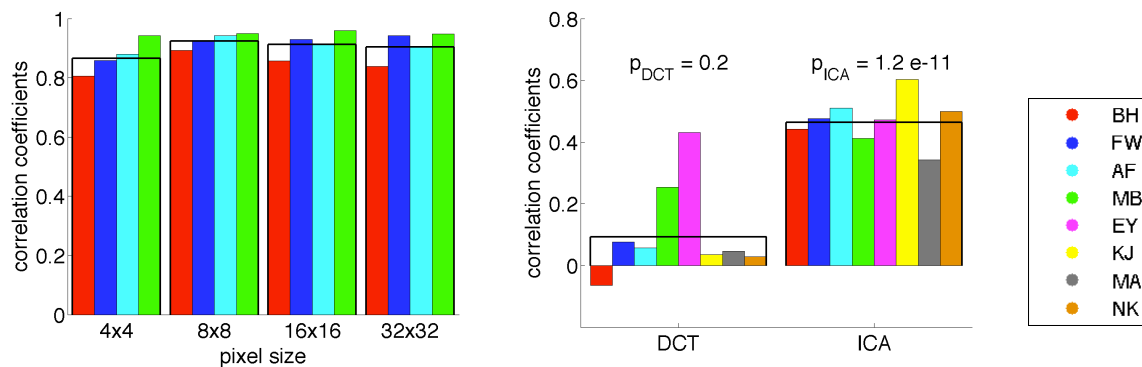


Figure 8. (Left) Bar charts show the correlation coefficients for different patch sizes of the pixel guessing-game. (Right) Shows the correlation coefficient between true and observer-set DCT- and ICA-basis coefficients, respectively. Data for individual observers shown as narrow coloured bars, mean across observers as wide, transparent bars with black contours.

function of ICA is more precisely to the point of the redundancy reduction hypothesis, it must not be forgotten that the search space of the optimization in ICA is still limited to linear transformations. Therefore only those redundancies can be removed which are linearly predictable.

In fact, it is known that the coefficients of the ICA basis for natural images are not truly independent^{31,32}. Moreover, it has been shown recently that the additional gain obtained by the optimization for higher-order redundancies in comparison with decorrelation methods is only 5% or smaller²⁵. The result of the present work goes beyond those theoretical considerations as it confirms empirically that the residual redundancies in the ICA coefficients are in fact of perceptual relevance.

This also suggests how to answer the other question: How is it possible that the DCT basis exhibited no significant correlations while the ICA basis did? The answer must be that the statistical dependencies between the DCT coefficients are perceptually irrelevant as they get attenuated during the process of perception.

In conclusion, our empirical findings suggest that the statistical redundancies between the ICA coefficients are perceptually more important than those of the DCT basis. Thus, perceptual dependence cannot merely be seen as a monotonic function of statistical dependence—there are many open questions both theoretical and psychophysical^{9,33-36} before we can claim to have understood early vision.

ACKNOWLEDGMENTS

We would like to thank Bruce Henning and Jakob Macke for their help with the manuscript and the Max Planck Society for financial support. This work is also supported by the German Ministry of Education, Science, Research and Technology through the Bernstein award (BMBF; FKZ: 01GQ0601).

REFERENCES

1. D. Kersten, "Predictability and redundancy of natural images," *Journal of the Optical Society of America A* **4**(12), pp. 2395–2400, 1987.
2. K. Koffka, *Principles of Gestalt Psychology*, Harcourt, Brace & World, New York, 1935.
3. S. W. Kuffler, "Discharge patterns and functional organization of mammalian retina," *Journal of Neurophysiology* **16**, pp. 37–68, 1953.
4. D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the animal, -cat's visual cortex," *Journal of Physiology (London)* **148**, pp. 574–591, 1962.
5. C. Enroth-Cugell and J. G. Robson, "The contrast sensitivity of retinal ganglion cells of the cat," *Journal of Physiology (London)* **187**, pp. 517–522, 1966.

6. D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *Journal of Physiology (London)* **195**, pp. 215–243, 1968.
7. F. W. Campbell and J. G. Robson, "Application of fourier analysis to the visibility of gratings," *Journal of Physiology (London)* **197**, pp. 551–566, 1968.
8. C. Blakemore and F. W. Campbell, "On the existence of neurons in the human visual system selective to the orientation and size of retinal images," *Journal of Physiology (London)* **203**, pp. 237–260, 1969.
9. G. B. Henning, B. G. Hertz, and D. E. Broadbent, "Some experiments bearing on the hypothesis that the visual system analyzes patterns in independent bands of spatial frequency," *Vision Research* **15**, pp. 887–899, 1975.
10. E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," in *Computational models of visual processing*, M. S. Landy and J. A. Movshon, eds., pp. 3–20, MIT Press, Cambridge, MA, 1991.
11. H. Simon, *The Sciences of the Artificial*, MIT Press, 1968.
12. B. Dubuc and S. Zucker, "Complexity, confusion, and perceptual grouping part i: The curve-like representation," *International Journal of Computer Vision* **42**(1/2), pp. 55–82, 2001.
13. F. Attneave, "Some informational aspects of visual perception," *Psychological Review* **61**, pp. 183–193, 1954.
14. H. B. Barlow, "Possible principles underlying the transformation of sensory messages," in *Sensory Communication*, W. A. Rosenblith, ed., pp. 217–234, MIT Press, Cambridge, MA, 1961.
15. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America A* **4**, pp. 2379–2394, 1987.
16. J. H. van Hateren, "A theory of maximizing sensory information," *Biological Cybernetics* **68**(1), pp. 23–29, 1992.
17. J. J. Atick, "Could information theory provide an ecological theory of sensory processing?," *Network: Computation in Neural Systems* **3**, pp. 213–251, 1992.
18. J. H. van Hateren, "Spatiotemporal contrast sensitivity of early vision," *Vision Research* **33**(2), pp. 257–267, 1993.
19. D. J. Field, "What is the goal of sensory coding?," *Neural Computation* **6**, pp. 559–601, 1994.
20. B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature* **381**, pp. 607–609, 1996.
21. B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1?," *Vision Research* **37**(23), pp. 3311–3325, 1997.
22. E. P. Simoncelli and B. A. Olshausen, "Natural images statistics and neural representation," *Annual Review of Neuroscience* **24**, pp. 1193–1215, 2001.
23. J. H. Van Hateren and A. Van Der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proceedings of the Royal Society of London Series B-Biological Sciences* **265**, pp. 359–366, 1998.
24. J. Atick, "Could information theory provide an ecological theory of sensory processing?," *Network* **3**, pp. 213–251, 1992.
25. M. Bethge, "Factorial coding of natural images: How effective are linear models in removing higher-order dependencies?," *Journal of the Optical Society of America A* **23**(6), pp. 1253–1268, 2006.
26. A. J. Bell and T. J. Sejnowski, "The "independent components" of natural scenes are edge filters," *Vision Research* **37**(23), pp. 3327–3338, 1997.
27. J. H. van Hateren and D. L. Ruderman, "Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex," *Proceedings of the Royal Society of London B* **265**, pp. 2315–2320, 1998.
28. A. Hyvärinen and P. O. Hoyer, "Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces," *Neural Computation* **12**(7), pp. 1705–1720, 2000.
29. A. Hyvärinen and P. O. Hoyer, "A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images," *Vision Research* **41**, pp. 2413–2423, 2001.
30. C. Shannon, "Prediction and entropy of printed english," *Bell. Syst. Tech. J.* **30**, pp. 50–64, 1951.

31. C. Zetsche, B. Wegmann, and E. Barth, "Nonlinear aspects of primary vision: entropy reduction beyond decorrelation," in *Int'l Symposium, Soc. for Information Display*, **XXIV**, pp. 933–936, 1993.
32. E. Simoncelli, "Statistical models for images: compression, restoration, and synthesis," in *31st Asilomar Conference on Signals Systems, and Computers Pacific Grove, CA*, pp. 673–678, IEEE Computer Society, Los Alamitos, CA, 1997.
33. A. M. Derrington and G. B. Henning, "Some observations on the masking effects of two-dimensional stimuli," *Vision Research* **29**(2), pp. 241–246, 1989.
34. G. B. Henning, C. M. Bird, and F. A. Wichmann, "Contrast discrimination with pulse-trains in pink noise," *Journal of the Optical Society of America A* **19**(7), pp. 1259–1266, 2002.
35. B. A. Olshausen and D. J. Field, "How close are we to understanding v1?," *Neural Comput* **17**(8), pp. 1665–1699, 2005.
36. G. B. Henning and F. A. Wichmann, "Some observations on the pedestal effect," *Journal of Vision* **7**(1), pp. 1–15, 2007.