NOVA

University of Newcastle Research Online

nova.newcastle.edu.au

1   Manipulating the structure of natural scenes using wavelets to study the

2   functional architecture of perceptual hierarchies in the brain

3

4   Alexander M. Puckett[a,b], Mark M. Schira[c], Zoey J. Isherwood[d], Jonathan D. Victor[e],

5   James A. Roberts[f], Michael Breakspear[g]

6

7   [a]School of Psychology, The University of Queensland; Brisbane QLD 4072; Australia

8   [b]Queensland Brain Institute, The University of Queensland; Brisbane QLD 4072; Australia

9   [c]School of Psychology, University of Wollongong; Wollongong NSW 2522; Australia

10   [d]School of Psychology, University of Nevada; Reno NV 89557; USA

11   [e]Feil Family Brain and Mind Research Institute and Department of Neurology, Weill Cornell Medical

12   College; New York NY 10065; USA

13   [f]Brain Modelling Group, QIMR Berghofer Medical Research Institute; Brisbane QLD 4006; Australia

14   [g]Brain and Mind PRC, University of Newcastle; Newcastle NSW 2308; Australia

15

16   ***Corresponding author:***

17   Alexander M. Puckett

18   School of Psychology

19   University of Queensland

20   St. Lucia QLD 4072

21   Australia

22   Email: pucketta@alumni.msoe.edu

23   Phone: +61435301757

Abstract

Functional neuroimaging experiments that employ naturalistic stimuli (natural scenes, films, spoken narratives) provide insights into cognitive function "in the wild". Natural stimuli typically possess crowded, spectrally dense, dynamic, and multimodal properties within a rich multiscale structure. However, when using natural stimuli, various challenges exist for creating parametric manipulations with tight experimental control. Here, we revisit the typical spectral composition and statistical dependences of natural scenes, which distinguish them from abstract stimuli. We then demonstrate how to selectively degrade subtle statistical dependences within specific spatial scales using the wavelet transform. Such manipulations leave basic features of the stimuli, such as luminance and contrast, intact. Using functional neuroimaging of human participants viewing degraded natural images, we demonstrate that cortical responses at different levels of the visual hierarchy are differentially sensitive to subtle statistical dependences in natural images. This demonstration supports the notion that perceptual systems in the brain are optimally tuned to the complex statistical properties of the natural world. The code to undertake these stimulus manipulations, and their natural extension to dynamic natural scenes (films), is freely available.

1. Introduction

Although the entire possible set of images that could be constructed (or imagined) is incredibly vast, the actual set of images encountered in the natural environment represents but a small subset of these possibilities (Field, 1994). All natural images share a number of characteristics, and this restricts the degree to which natural images occupy the state-space of all possible images. For example, the intensities, colors, and spectral properties of adjacent regions of a natural image are similar – with the correlation decreasing with distance (Burton and Moorhead, 1987; Frazor and Geisler, 2006). This lower-order pattern of pairwise correlations is, however, only part of the picture. Natural images also share a number of higher-order statistical relationships (Graham et al., 2016; Hermundstad et al., 2014; Karklin and Lewicki, 2009; Tkacik et al., 2010). For example, spectral properties at one spatial scale (such as high contrast edges) are conditionally dependent on those at other scales (such as shading and contours). Together, these statistical properties impart the spatial structure typical of natural images – that is, they produce the patterns we associate with trees, forests, faces, rivers, rocks, and the like.

Given that all natural images are structured in a statistically similar way, it is not surprising that the mammalian visual system appears to be specifically tuned for this structure. A great deal of work has been done to elucidate the response properties of neurons in the visual cortex of a number of mammals (e.g., cat, monkey, and man) (Hubel and Wiesel, 1959, 1968; Yoshor et al., 2007). Across these species, it has been shown that the receptive fields in primary visual cortex are spatially localized, oriented, and selective to structure at various spatial scales (i.e., acting as bandpass filters) (Field, 1999). It has been suggested that, by being sensitive to specific spatial

70   frequencies and orientations, the simple cells in primary visual cortex are matched to

71   the higher-order structure found in natural images. Pertinently, it has been shown that

72   filters modeled after these simple cells (i.e., similar orientation and bandpass

73   parameters) respond with a high degree of kurtosis when presented with images of

74   natural scenes. That is, they respond particularly precisely to local features in natural

75   scenes with properties matched to their preferred stimulus properties. Moreover, this

76   kurtosis diminishes when the filter parameters differ from those found in the

77   mammalian visual system (Sekuler and Bennett, 2001) so that they respond less

78   precisely and more diffusively to local stimulus features. This has been interpreted as

79   evidence that the visual system is developed to optimize the coding of natural image

80   content as the high degree of kurtosis leads to sparse, distributed responses – an

81   efficient coding strategy whereby most of the information for each instance of a specific

82   natural scene is represented by a small, unique set of cells (Field, 1999).

83

84   To account for such response properties of neurons in primary visual cortex and their

85   sparse coding of natural image content, it has been shown that receptive fields can be

86   represented mathematically by a wavelet-like transform. The wavelet transform is

87   similar to the more widely known Fourier transform in the sense that it can decompose

88   a very broad variety of functions and empirical data into a set of oscillatory basis

89   functions. However, rather than transforming the data into a domain of simple sine and

90   cosine functions, the wavelet transform represents the data with more complex

91   functions – called wavelets (Graps, 1995). These functions are localized in space and

92   process data at different spatial scales – similar to the receptive fields in mammalian

93   visual cortex. Importantly, whereas successive frequencies in the Fourier domain are

94   linearly spaced, successive wavelet scales are dyadic and hence logarithmically

95   spaced – that is, every scale is twice (or half) the frequency than the level above (or

96   below). Hence, when applied to images of natural scenes, different wavelet functions

97   are sensitive to the sparse, higher-order statistical structure that is present at different

98   spatial scales (Field, 1999; Olshausen and Field, 1996).

99

100   Understanding and manipulating the statistics of natural scenes holds potential to test

101   the hypothesis that the visual system is tuned to their expected (typical) properties.

102   Here we exploit the relationship between receptive field properties and wavelets to

103   manipulate the higher-order statistical structure in natural scenes. This paper

104   comprises two distinct but complementary parts. In the first part, we show how the

105   wavelet transform can be used to parametrically degrade natural image structure: (1)

106   at specific spatial scales, (2) in a global or locally-targeted fashion, and (3) for dynamic

107   (i.e., films) as well as static scenes. We first provide a didactic introduction to wavelet

108   resampling. We then provide novel extensions to adopt the procedure from its classic

109   application in non-parametric inference to its use in naturalistic paradigms, preserving

110   the color palette of stimuli, and manipulating dynamic natural scenes (films). We also

111   present a novel extension using incremental resampling to more deeply probe the

112   statistical structure of natural scenes and their relationship to other natural

113   phenomena. In the second part, we demonstrate the utility of this approach by showing

114   how it can be used to create stimuli that can be used along with fMRI to probe the

115   hierarchy of human visual cortex – showing that cortical responses at different levels

116   of the visual stream are differentially sensitive to the subtle, wavelet-based parametric

117   statistical manipulations.

118

119   **2. *Manipulating natural image structure – the wavelet transform***

120   Natural images are usually defined as any image of the natural, physical, or material

121   world and can portray general scenes (e.g., beaches, forests, mountain ranges) or

122   specific objects (e.g., rocks, trees, waterfalls). Figure 1A, a photograph of a patch of

123   fallen leaves, is an example of such a natural image. Contrasting this natural image

124   with luminance-matched noise images (Fig. 1B,C) provides insight into the structure

125   and properties of natural images. Figure 1B was generated by random assignment of

126   pixel luminance values from Figure 1A (i.e., white noise) and has little in common with

127   natural images. Figure 1C is also random but was generated with the additional

128   constraint that the distribution of energy across spatial frequencies matched that of

129   the natural image. That is, it is characterized by a similar $1/f^{\alpha}$ amplitude spectrum (Fig.

130   1D) — a property which describes the distribution of amplitude (luminance intensity)

131   as a function of spatial frequency. Across natural scenes, the slope (α) of this

132   distribution is remarkably similar with values typically ranging between 0.8-1.2 (Burton

133   and Moorhead, 1987; Field, 1987; Ruderman and Bialek, 1994; Tolhurst et al., 1992;

134   van der Schaaf and van Hateren, 1996). If the distribution of luminance intensity

135   variations in nature was random and independent of spatial scale, then natural scenes

136   would possess the amplitude spectra of white noise ($\alpha = 0$) (Fig. 1B), where amplitude

137   is the same across all spatial frequencies.

138

139   Despite the similarity between the amplitude spectra of an actual natural scene (Fig.

140   1A) and of "natural" (or colored) noise (Fig. 1C), one would have no trouble identifying

141   the true natural scene. This demonstrates how matching lower-order statistical

142   properties is insufficient to produce the structure present in natural images. Rather,

143   the structure is a consequence of higher-order statistical relationships. Being able to

144   parametrically manipulate these statistical dependences permits the controlled

145 investigation of how the visual system processes this structure and is the main

146 objective of the wavelet technique described below.
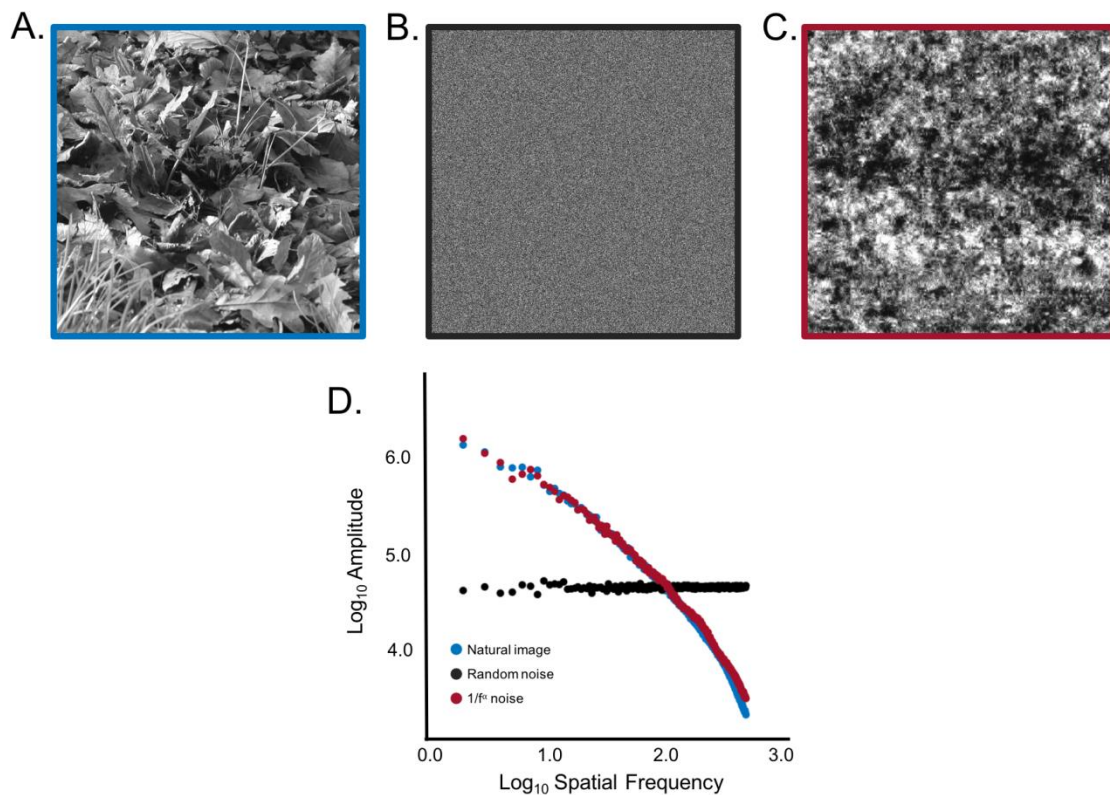
Figure 1. Difference between a natural image and noise. (A) Natural image. (B) Random noise. (C) $1/f^\alpha$ noise. (D) Spatial frequency spectra for A-C. Note that the image in A is from the Zurich natural images database (Einhauser and Konig, 2003).

152 To manipulate natural image structure using wavelets, the discrete wavelet transform

153 (DWT) is first used to perform a multi-resolution decomposition of the image data

154 (Breakspear et al., 2004). This decomposition uses a family of wavelet basis functions

155 sensitive to variance at specific spatial scales. At each scale, the data are

156 decomposed into two orthogonal components containing information about the

157 variation in signal intensity at that spatial scale (i.e., the detail coefficients) and the

158 residual of the signal after those and all smaller details have been removed (i.e., the

159 approximation coefficients). Because the image data is two-dimensional, the detail

160 coefficients are further decomposed into horizontal, vertical, and diagonal

161 components. Note that the original image can be recovered, without loss, by linearly

162 adding the approximation of the signal at a specific spatial scale together with the

163 details at that scale and all smaller scales. A more detailed description of the two-

164 dimensional DWT can be found in the Supplementary Material (S1).

165

166 2.1 Degrading scale-specific information

167 As emphasized above, the DWT yields a representation of the image data across a

168 hierarchy of spatial scales. Whereas the original image is spatially correlated, the DWT

169 is a "whitening" transform and adjacent wavelet coefficients are statistically

170 independent (Bullmore et al., 2001). It is therefore possible to randomly permute the

171 detail coefficients within any level of this hierarchy – essentially destroying the higher-

172 order statistical dependences at the specific spatial scale represented by that level

173 without loss of energy. This crucially differs from smoothing, filtering, or adding noise

174 to the data. Following this permutation, the inverse DWT is performed, yielding an

175 image nearly identical to the original but without structure at the targeted spatial scale.

176 Figure 2 illustrates the results of this process in which the structure present in a natural

177 image (Fig. 2A) is degraded at individual spatial scales (Fig. 2B,C) as well as at

178 multiple scales (Fig. 2D,E). Importantly, this process only degrades the higher-order

179 statistical relationships while maintaining the lower-level image content such as the

180 contrast, luminance histogram, and spatial frequency content (Fig. 2F).
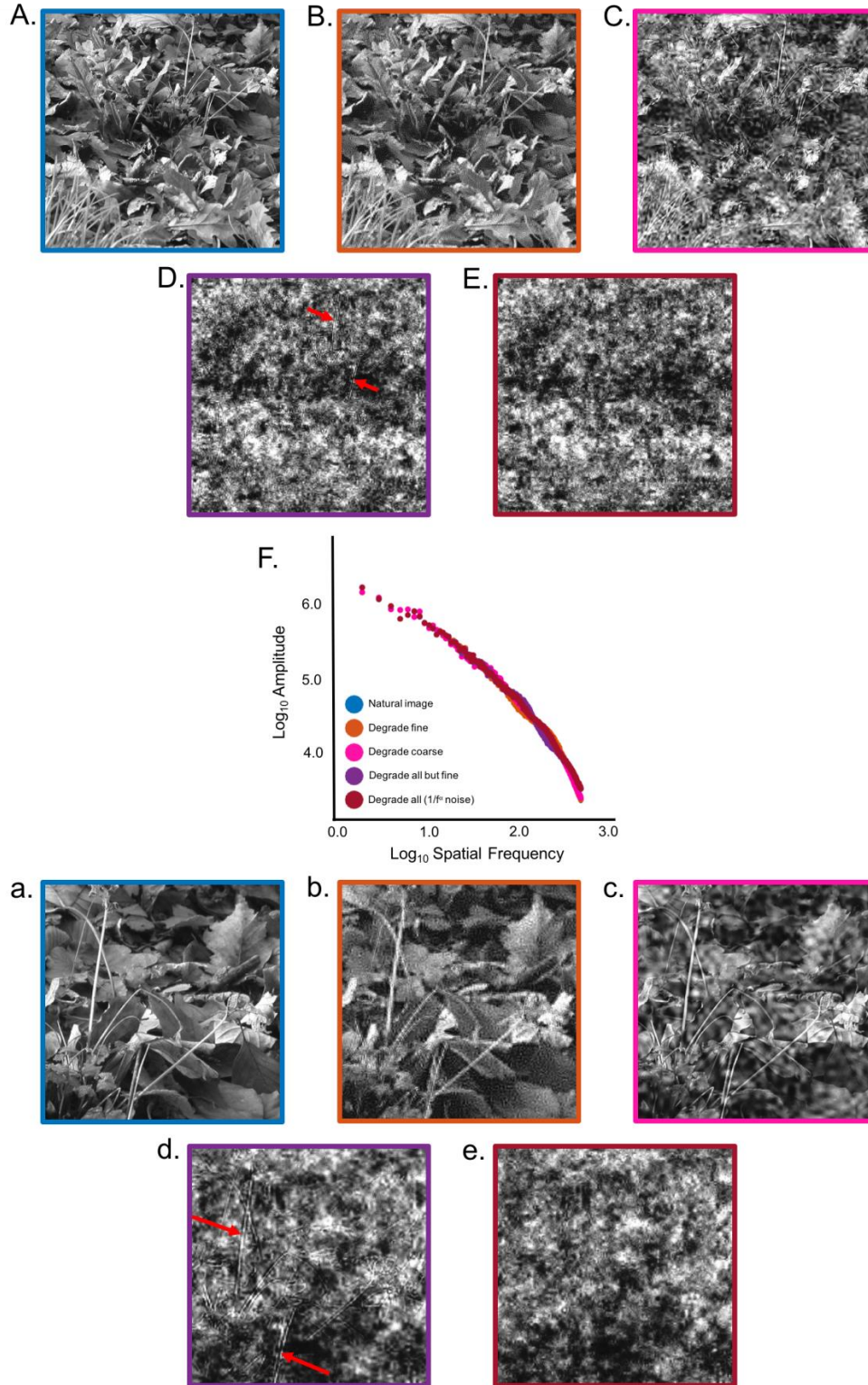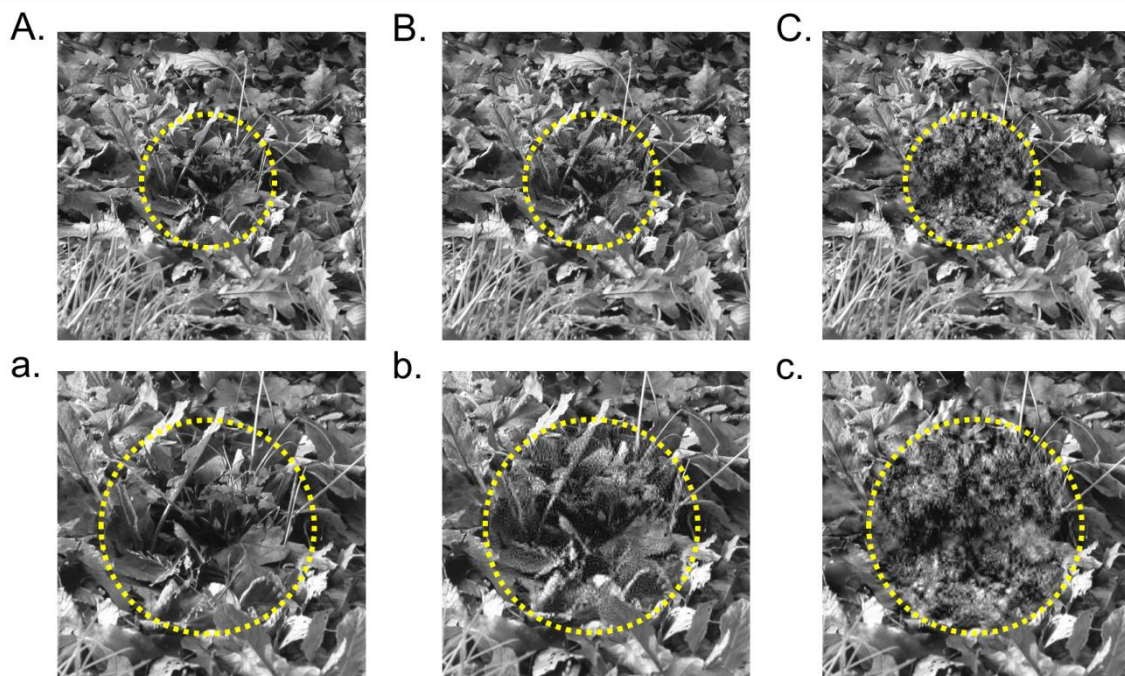
Figure 2. Using wavelets to degrade scale-specific natural image structure. (A) Intact natural image. (B) Natural image with fine scale structure degraded. (C) Natural image with coarse scale structure degraded. (D) Natural image with all scales of structure degraded except the fine scale (red arrows indicate examples of remaining fine scale structure). (E) Natural image with all scales of structure degraded (i.e.,$1/f^\alpha$ noise). (F) Spatial frequency spectra for A-E. Lowercase a-e show a zoomed-in view (upper-right quadrant only) of images A-E to aid observation of the manipulations.

189 Inspection of this process reveals the effects of degrading the structure present in a

190 natural image at various spatial scales. Close inspection of Figure 2B(b) reveals that

191 the very fine structures have been degraded – including veins of leaves and the sharp

192 edges of the plant blades. This is in contrast to Figure 2C(c) in which the finer details

193 are still present, but coarser structures (e.g., at the level of entire leaves) have been

194 disrupted. Figure 2D(d) illustrates the effect of degrading the structure at all scales

195 except the fine scale with the image being nearly devoid of all natural image structure.

196 However, from what is otherwise a pure colored noise image, one can distinctly make

197 out the very sharp edge details that were otherwise degraded in Figure 2B(b). Finally,

198 Figure 2E(e) illustrates the effect of degrading this remaining scale of information

199 (along with all others) – producing a colored noise image with no apparent natural

200 image structure but with nearly identical low-level image content as the original natural

201 image (Fig. 2F). That is, the original and wavelet scrambled (or "wavestrapped") data

202 are essentially identical in terms of very basic visual features (e.g., luminance,

203 contrast, and spectral content). The more elusive properties that couple details, edges,

204 and outlines to depth, shadows, and context – and that convey the meaningful

205 properties of natural visual scenes – have been randomized.

206

207 2.2 Wavestrapping can be spatially-localized.

208 Unlike the Fourier transform, the wavelet basis functions are localized in space. This

209 attribute makes it possible to use the wavelet transform to degrade natural image

210 structure in a spatially-restricted manner, rather than uniformly across the entire

211 image. The procedure is similar to that described above, except that only detail

212 coefficients associated with a specific spatial domain are permuted before performing

213 the inverse DWT – the detail coefficients outside that domain are left unchanged. The

214 result of such a spatially-restricted degradation are illustrated in Figure 3. Here, we

215 have independently resampled the coefficients associated with the central region of a

216 natural scene image and its surround. If fixating at the centre of the image, this

217 procedure can be used to degrade natural image structure to probe foveal vs.

218 peripheral visual processing. Notably, any spatial domain can be used to restrict the

219 permutation process. This same basic procedure can hence be used to target

220 processing associated with specific hemifields or quadrants of an image.

221

222

223 **Figure 3.** Using wavelets to degrade a spatially-restricted area. (A) Intact natural image with yellow
224 dashed circle denoting the targeted foveal region. (B) Natural image with only fine scale structure
225 degraded near the fovea. (C) Natural image with all structure degraded near the fovea. Lowercase a-c
226 show zoomed-in views of the central regions in A-C.
227

228 2.3. Extension to color images

229 The wavestrapping approach can be extended to color images (Fig. 4A). However,

230 the addition of color information does require further considerations. While each pixel

231 in a grayscale image can be described by a single number (intensity), color images

11

232 contain three numbers per pixel – one for each color channel: red, green, and blue.

233 The simplest extension of the above randomization techniques to a color image is to

234 degrade the spatial structure in each channel independently. However, doing so does

235 not preserve the color palette (Fig. 4B). To preserve the original colors (the color

236 equivalent of preserving the pixel amplitude distribution), the image structure within

237 each channel needs to be permuted in the same way across channels. In practice this

238 can be achieved by permuting the detail coefficients within each color channel

239 beginning with the same random seed (Fig. 4C).

240
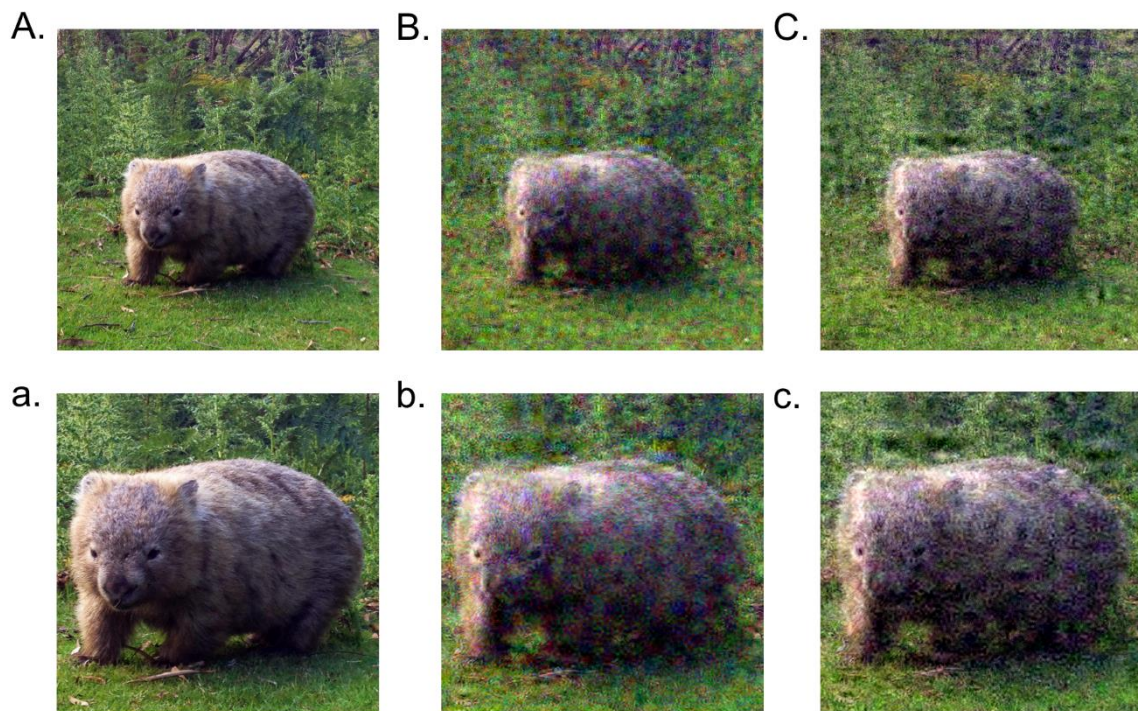
241
242 **Figure 4.** Application to color images. (A) Intact natural image with RGB color channels. (B) Image with
243 color channels degraded independently. (C) Image with color channels degraded identically. Note that
244 the color palette is preserved in C but not B. This can most easily be seen by the examining the body
245 of the wombat, which is tannish/brown in both A and C but mottled with red, green, and blue patches in
246 B. Lowercase a-c show a zoomed-in view of the images for closer examination. Source photo from
247 author A.M.P.
248

249   <u>2.4 Extension to naturalistic movies</u>

250   The above principles can be extended to dynamic natural scenes – i.e., film stimuli. In

251   this case there is the additional dimension of time. Film stimuli incorporate the rich

252   temporal variations in our environment and hence can provide a more engaging and

253   ecologically-valid naturalistic experience than traditional static images (Hasson et al.,

254   2004; Roberts et al., 2013; Sonkusare et al., 2019). The key consideration then is how

255   to handle the temporal domain alongside the degradation of the spatial dimensions.

256   One simple possibility is to permute the (spatial) wavelet coefficients within each frame

257   independently, breaking the temporal structure associated with the scrambled spatial

258   scales. However, this whitens the temporal spectra – introducing spurious high

259   frequencies – as each frame differs abruptly from the preceding one. To fully preserve

260   the temporal structure, one can use the same random seed for each frame (and for

261   color videos, within each color channel too). Even with all spatial scales scrambled,

262   preserving the temporal structure leaves an "imprint" of moving objects within the

263   scene, as well as pans and cuts (see Supplementary Material S2, Sup Movie 1 for an

264   example). Given the importance of motion to the visual system – including the

265   "biological motion" of humans (Allison et al., 2000; Schultz and Pilz, 2009) – this

266   preservation of apparent motion is crucial when permuting dynamic films in the wavelet

267   domain to study the visual cortex.

268

269   This second strategy of a constant random seed destroys higher order statistics in the

270   spatial domain but leaves those in the temporal domain exactly preserved. Wavelet

271   resampling can also be applied in the temporal domain, treating the video as a single

272   multidimensional time series, rather than as a series of discrete two-dimensional

273   images. Notably, temporal variance of dynamic natural scenes also possesses a $1/f^{\alpha}$

274 amplitude spectrum (Fig. 5A). This spatio-temporal wavestrapping can be achieved in

275 two steps: parallel two-dimensional spatial resampling followed by parallel one-

276 dimensional temporal resampling (Fig. 5B). Alternatively, the entire film could be

277 wavestrapped using a single three-dimensional DWT following the same principles as

278 wavestrapping a single three-dimensional spatial object (such as a single whole-brain

279 fMRI volume (Breakspear et al., 2004)), although this mixes together information from

280 the spatial and temporal domains.

281

282 Using wavelets to manipulate movie data in the time domain can also adopt extensions

283 outlined above for spatial images – namely focusing on high or low temporal scales

284 and/or choosing specific temporal moments (such as scene transitions) and leaving

285 other blocks unchanged. Temporal resampling can also extend to the parallel stream

286 of audio information.

287
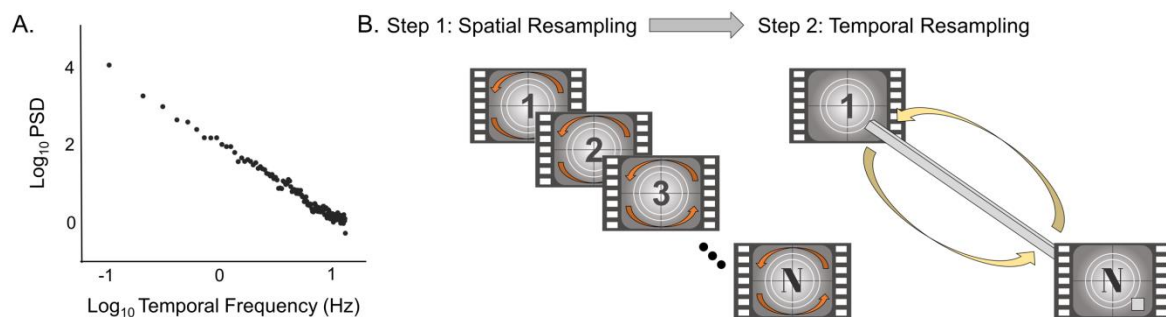
288
289 **Figure 5.** Extension of wavestrapping to movies. (A) Temporal spectrum from a film clip shown as the
290 power spectral density (PSD) across temporal frequency. Note that the spectrum was calculated from
291 the red channel, middle pixel of Supplementary Movie 1 using a 10 second window and 50% overlap.
292 (B) Schema for two-step wavestrapping of films. In Step 1, each frame and at each time point is spatially
293 resampled (indicated by orange arrows). The resampling procedure is identical at the same scale for
294 each time point and each frame. In Step 2, the time series from each pixel from the spatially
295 wavestrapped data is resampled in the temporal dimension (indicated by yellow arrows). The
296 resampling procedure at the same scale for each voxel is identical. All resampling is performed in the
297 wavelet domain after appropriate wavelet decomposition (two-dimensional for Step 1 and one-
298 dimensional for Step 2).
299

2.5 Thermodynamics of natural scenes

Recent work has shown that static (Saremi and Sejnowski, 2013) and dynamic (Munn and Gong, 2018) natural scenes possess the statistical hallmarks of criticality – that is, they reside close to a phase transition (i.e., a statistical boundary) between order and disorder. Computational analyses of natural scenes using the methods of statistical mechanics has suggested that this phase transition resides within specific latent layers of a natural scene (Saremi and Sejnowski, 2013) and is associated with thermodynamic "frustration" (see Supplementary Material, S3). By residing near these phase transitions, natural scenes are able to reflect a critical balance between (1) the ordered arrangement of the contours, edges, and textures of various sizes that endow it with structure and information and (2) the idiosyncratic and haphazard nature of this arrangement into the objects that characterize any specific scene and hence yield its semantic meaning and unique visual impression.

By applying our wavestrapping approach progressively it is possible to demonstrate the balance between order and disorder inherent to natural images (Figure 6). This is because the randomization can be realized in varying degrees of depth, from just a few permuted coefficients up to full permutation. This is achieved by selecting random subsets of coefficients for permutation, leaving others invariant. Figure 7A and Movie 1 both demonstrate the process of progressively disordering a natural image, which can be thought of as "heating" the scene. As can be seen in Figure 7B, the amount of variability between realizations increases monotonically with the depth of randomization. Note that fully randomized realizations (i.e., randomization depth of 100%) are the most highly variable – akin to a gas. These highly variable realizations can be appreciated if one "boils" the scene (i.e., continues to randomize at a depth of

15

325   100% - see Movie 2). However, incremental permutations do show scale-specific

326   expressions of variability which differ between scenes (Fig. 7C). That is, despite their

327   featureless 1/f spatial spectra, each natural scene has a distinct signature of

328   increasing variability at different scales. Incremental wavelet resampling thus unpacks

329   the latent statistical frustration within natural scenes which is not uniform across scales

330   and scenes.

331
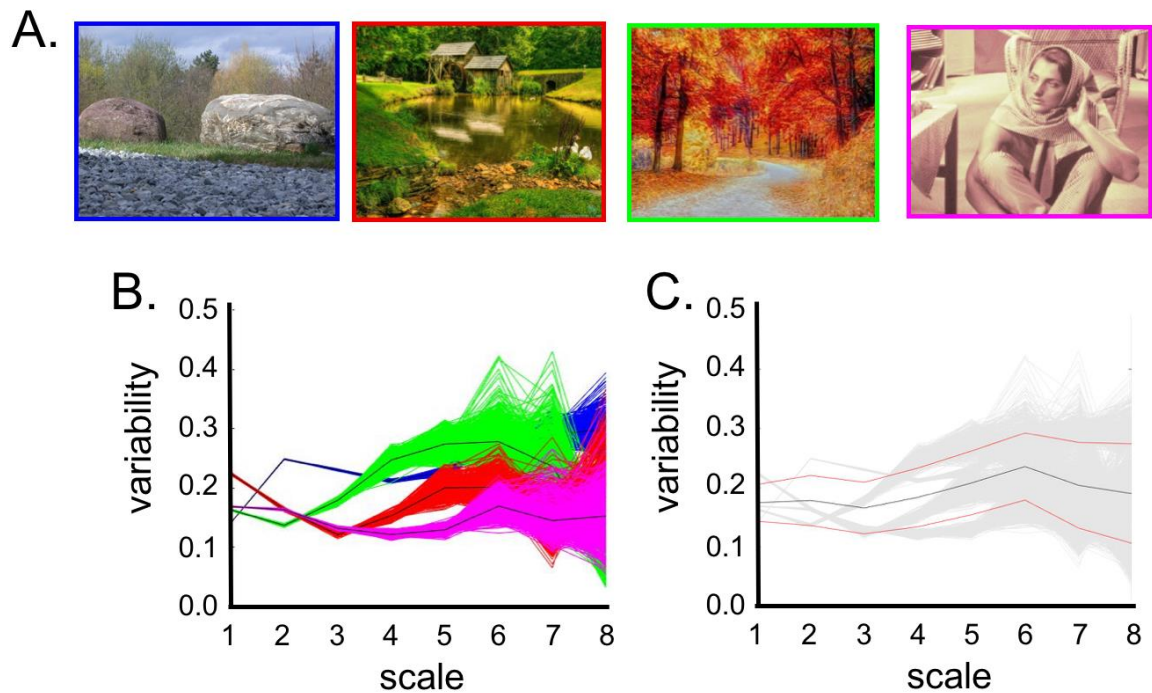
332
333 **Figure 6**: (A) Images used to demonstrate the thermodynamic properties of natural scenes. (B)
334 Variability vs. wavelet scales. Each colored line is a single permutation of the corresponding image at
335 one scale. Image variability is measured as the root mean squared differences between the original
336 and scrambled image across pixels. Black lines show image averages. (C) Mean (black) across all four
337 images ± standard deviation (red). There are no trends in mean image variability.
338
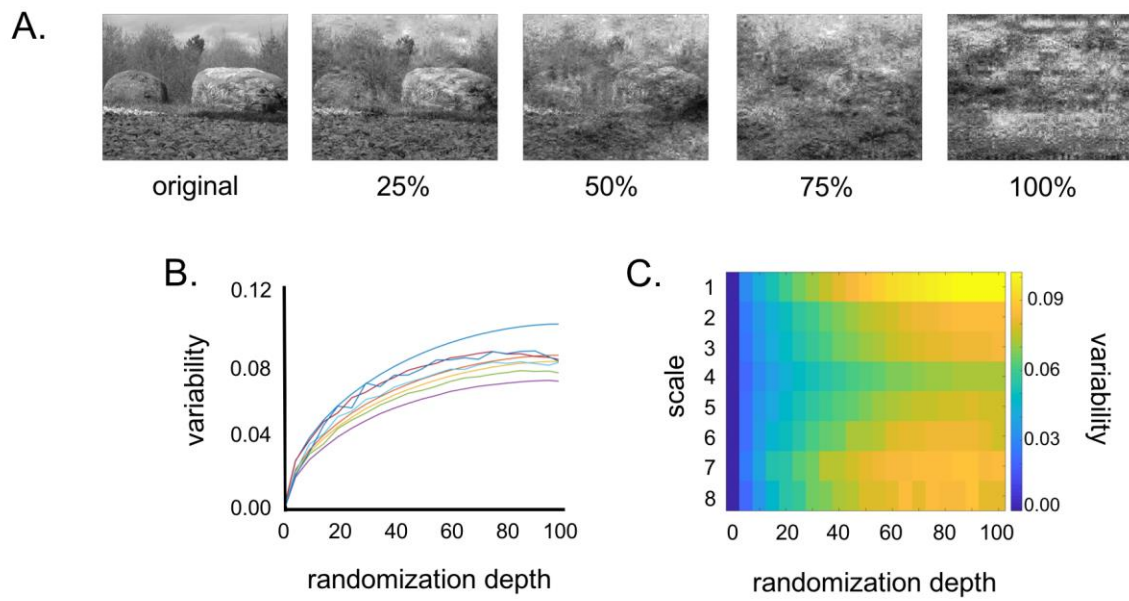
Figure 7: (A) Wavelet-based randomization ("heating") of a natural scene, increasing incrementally from the original scene to fully randomized in steps of 25%. (B) Variability amongst an ensemble of random realizations increases monotonically with increasing depth of randomization at all scales. (C) However, some scales (here fine and coarse) show slightly greater variability with randomization than others (here mid-scales).

Movie 1: Wavelet-based randomization of a natural scene, increasing incrementally from the original scene to fully randomized. This process is akin to "heating" the scene.

Movie 2: Wavelet-based randomization of a natural scene, iteratively applied full randomization. This process is akin to constant "boiling" of the scene.

The wavelet-based randomization (or heating) can easily be reversed. For example, Movie 3 shows the process of "cooling" the scene back down from a boil (i.e., a fully randomized state) to its natural state. Interestingly, we can then continue to cool the image beyond its natural state and hence approach a single ordered state – akin to a solid (Movie 4). This process of "freezing" is further demonstrated in Figure 8A, which shows the process of progressive ordering of a natural image. As can be seen in Figure 8B, the amount of variability between realizations increases to a maximum at approximately 50% of wavelets ordered, corresponding to a mixture of natural and ordered phases, then decreases again as the single ordered state is approached. Similar to the process of randomization, the incremental ordering permutations do show scale-specific expressions of variability which differ between scenes (Fig. 8C). Finally, we can "thaw" a frozen image (i.e., in a 100% ordered state) to its original, natural state (Movie 5) by progressive randomization (or heating) as described above.

366

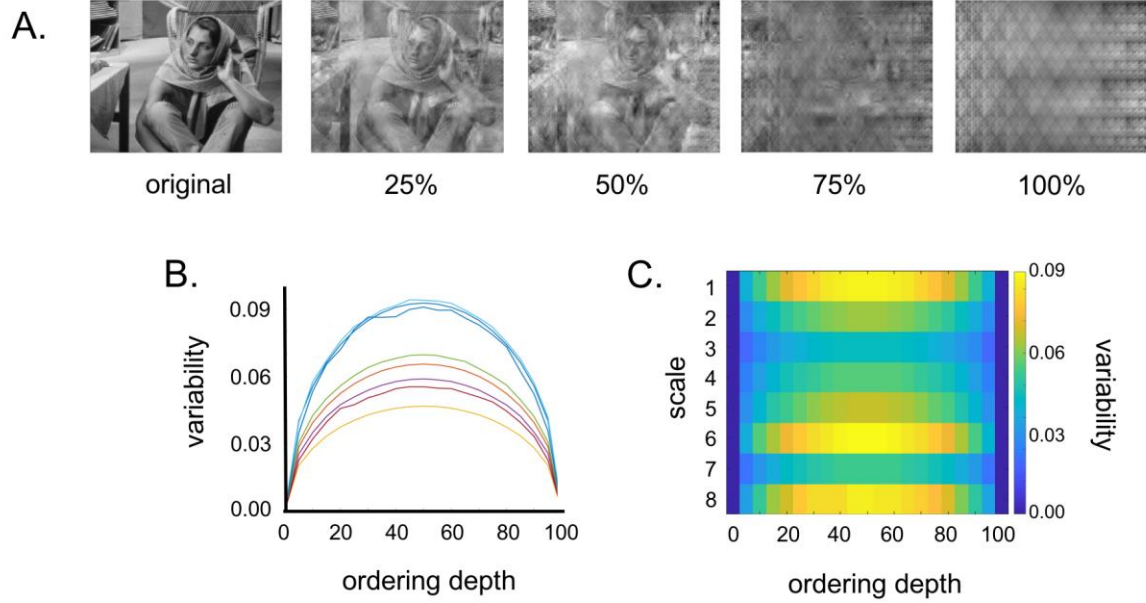

367
368 **Movie 3:** Wavelet-based ordering of a natural scene, decreasing incrementally from the fully
369 randomized state to the original scene. This process is akin to "cooling" the scene.
370
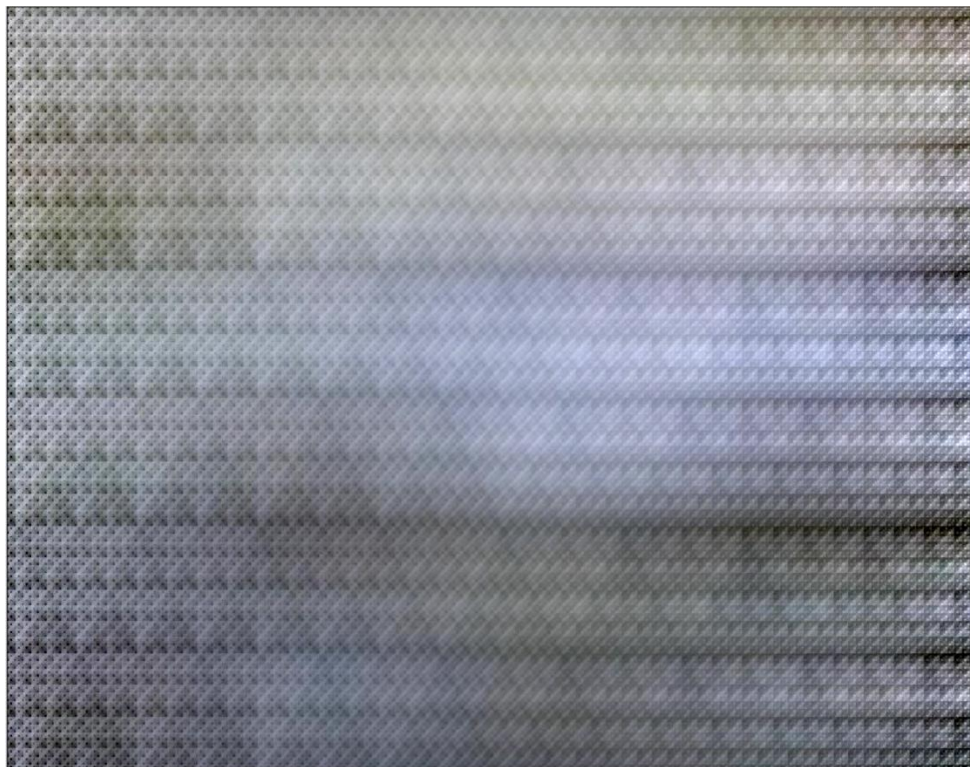
371
372 **Movie 4:** Wavelet-based ordering of a natural scene, decreasing incrementally from the original scene
373 to a single ordered (or "frozen") state.
374

Figure 8: (A) Wavelet-based ordering ("cooling") of a natural scene, increasing incrementally from the original scene to fully ordered in steps of 25%. (B) Variability amongst an ensemble of realizations increases to a maximum at approximately 50% of wavelets ordered, corresponding to a mixture of natural and ordered phases, then decreases again as the single ordered state is approached. (C) Some scales show greater variability with ordering than others.

Movie 5: Wavelet-based randomization of a natural scene, decreasing incrementally from the single ordered (or "frozen") state back to the original scene.

Figure 9 further demonstrates the notion of natural image thermodynamics with natural images being positioned at a critical phase between fully ordered and disordered states. Subtle manipulations of dynamic natural scenes, using wavelet resampling to parametrically disrupt the complex statistics of their criticality whilst measuring cortical dynamics, represents an elusive but untested means of understanding how the structure of cortical dynamics are tuned adaptively to those of the natural world. Interestingly, the "critical" nature of dynamic natural scenes (i.e., that they are perched between order and disorder reflecting the balance of scene stability and sudden, spontaneous transitions) mirrors the critical, avalanche-like dynamics that occur throughout cortical systems (Cocchi et al., 2017). Incremental disruption – both "heating" (randomizing) and "cooling" (ordering) – allows tuning of a natural scene through its critical point and could be used in conjunction with imaging or neurophysiological recordings to further explore this intriguing area.
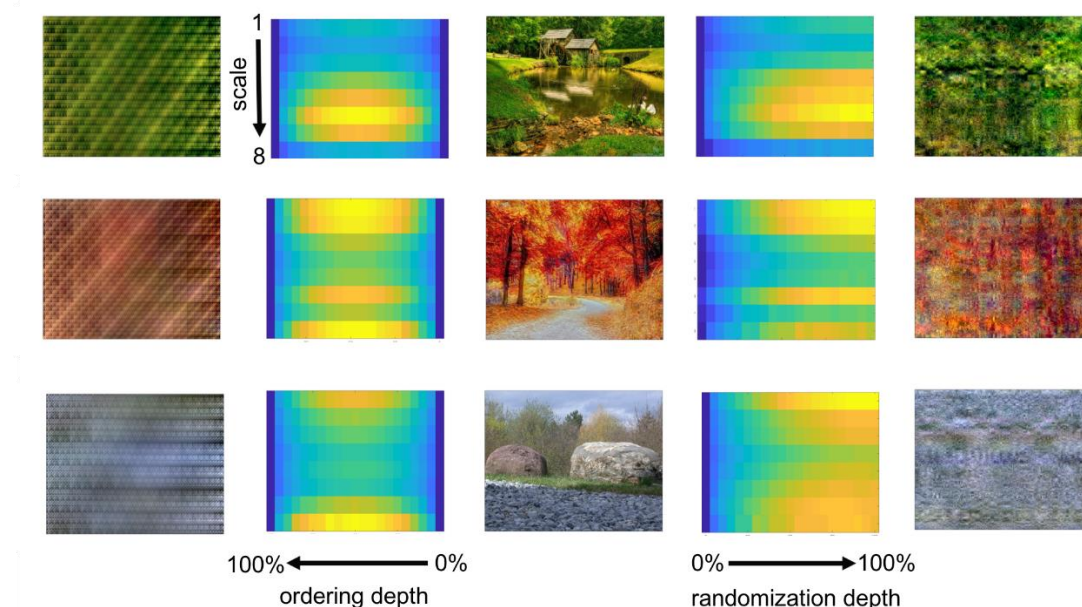
Figure 9: Natural images (middle column) reside near a critical boundary between order and disorder. Incremental, wavelet-based randomization (or heating) and ordering (or cooling) lead to fully disordered (right most column) vs. fully ordered states (left most column), respectively. Sandwiched between the images are plots of the variability seen across both scale and the depth of ordering or randomization when cooling or heating the natural image.

3. Probing the visual hierarchy – an fMRI demonstration

We conducted an fMRI experiment to illustrate the application of wavelet-based manipulations of natural images to probe the functional architecture of the visual hierarchy. As outlined above, there are numerous potential ways to manipulate static and dynamic natural scenes using wavelets. We designed a parametric, passive-fixation task to demonstrate some of the practical considerations of performing an fMRI experiment using wavelet-degraded stimuli (e.g., number of conditions can multiply quickly, use of a fixation task aimed at controlling attentional resources, etc.). Our proof-of-principle application to a visual fMRI experiment builds upon prior research in this field with the overarching goal being to contrast levels of cortical activity in different visual regions elicited by the presentation of intact natural images vs. wavelet-degraded natural images. Importantly, the basic image properties (luminance, spectra) remain the same between the two image types; only the higher-order statistical dependences (i.e., the structure of that image content) differ. To control for possible transition effects between (natural and wavestrapped) stimuli, we designed a factorial experiment which counterbalances the nature and order of their presentation.

Although primarily demonstrative, the experiment was motivated by a central hypothesis: that higher visual areas would be more sensitive to the complex structure present in natural images than lower visual areas. This was motivated by decades of previous research showing that primate visual cortex is organized hierarchically, with neurons responding to increasingly complex features as one progresses up the cortical hierarchy (DeYoe and Van Essen, 1988; Felleman and Van Essen, 1991; Van Essen, 2004).

432

3.1 Materials and Methods

3.1.1 Subjects

Seven, right-handed participants (22-24 years, mean 22.9 years; 3 male, 4 female) who disavowed a history of neurological or psychiatric diseases completed a functional neuroimaging experiment. All participants had normal or corrected to normal vision. The experiment was conducted with the written consent of each participant following approval by the local human research ethics committee in accordance with national guidelines.

3.1.2 Experimental design

Stimuli were presented in blocks of 8 seconds while participants fixated on a small superimposed crosshair. Stimuli consisted of natural images, degraded images obtained through wavestrapping these natural images at select (fine or coarse) scales, and colored noise control images matched for luminance and spectra content obtained through wavestrapping the natural images at all spatial scales.

A partial 3x2x2 within-subjects factorial design was used. The independent variables were type of image manipulation (N1: degrade from natural image, N2: degrade from noise, N3: restore from noise), spatial scale manipulated (S1: fine and S2: coarse), and presentation of manipulation (F1: flip vs. F2: flick). All experimental conditions are summarized in the Supplementary Material (S2, Table 1), with representative conditions described in detail below:

- N1,S1,F1 – the fine scale information (S1) of a natural image was permuted (N1). This resulted in the degradation of the structure at that scale and hence

457   a natural image with all scales of structure intact except the fine scale. The

458   experimental block involved flipping back and forth (F1) between the original

459   image and the degraded image.

460   • N1,S2,F2 – the coarse scale information (S2) of a natural image was permuted

461   (N1). This resulted in the degradation of the structure at that scale and hence

462   a natural image with all scales of structure intact except the coarse scale. The

463   experimental block involved flicking through (F2) a series permutations of the

464   same source image (i.e., the permutation was carried out a number of times on

465   the same natural image, and were presented in succession during the imaging

466   block).

467   • N2,S2,F2 – the coarse scale information (S2) of a noise image was permuted

468   (N2). Since we began with a noise image, there was no natural scene structure

469   to degrade; however, the permutation was identical to what was performed on

470   a natural image and leads to a slightly distinct noise image that differs only at

471   the targeted spatial scale. The experimental block involved flicking through (F2)

472   a series of permutations on the same source image.

473   • N3,S2,F1 – the coarse scale information (S2) from a natural image was put into

474   a noise image (N3). The experimental block involved flipping back and forth

475   (F1) between the original noise image and the noise image with structure

476   added.

477

478 These stimuli permutations were designed to parametrically control the depth of image

479 manipulation and the spatial scale targeted while controlling for the effects of image

480 transitions. The factorial design was incomplete (partial) in that it was not possible to

481 test the flick presentation type (F2) for the condition that involved adding structure to

a noise image (N3). That is, for any given natural scene there is only one possible instance of structure that can be added to remain faithful to the original scene (i.e., any alteration of this structure would change the scene). In contrast, there is no limit to the number of instances of noise images that can be constructed from each natural scene due to the randomized nature of the wavestrapped permutations. In addition to the above conditions (all of which involve a changing stimulus, whether flipping or flicking), we also included two static image block types: a natural image (N1, S0, F0) and a noise image (N2, S0, F0). An isoluminant grey background was shown as a baseline block.

In total then, there were 12 different experimental block types and a baseline. Each block was presented three times per scan run. All experimental blocks were presented for 8 seconds and the grey background baseline was presented for 12 seconds. During the ON period for the stimulus blocks with image change (i.e., flip or flick), the transition occurred every 0.5 seconds. The block types were pseudo-randomized except that we ensured that each block type followed the grey background baseline condition an equal number of times and that the last block of every run was the grey background condition to permit the fMRI signal to return to baseline. 12 runs were collected per subject, in a single scan session.

To control attention, aid fixation, and monitor subject alertness, a color/orientation conjunction task was performed at fixation throughout the entire run (Puckett and DeYoe, 2015; Treisman and Gelade, 1980). For this purpose, a small circle (10 x 10 pixels, subtending 0.15° visual angle) was superimposed upon the images. The circle contained a pattern that randomly changed every 2 seconds among 4 possible

configurations: red horizontal, red vertical, green horizontal, and green vertical. The

participant was required to report the nature of each change via one of two button

presses (button 1 = red horizontal or green vertical, button 2 = red vertical or green

horizontal). In addition to the color/orientation patch, a fine grid was overlaid on the

images to aid fixation (Schira et al., 2007).

An example of the visual stimulus and block paradigm (with annotation), is presented

in the Supplementary Material (S2, Sup Movie 2).

3.1.3 Image manipulation

Stimuli were constructed by manipulating a set of natural images using the wavelet

transform (as outlined in section 2). The natural images were sourced from the "Zurich

natural images" database, which is freely available for academic use (Einhauser and

Konig, 2003). Note that the subset of images from this database used here are shown

in the Supplementary Material (S2, Sup Figures 1 and 2). In general, constructing the

stimuli involved: converting the RGB image to greyscale, permuting the detail

coefficients at a specific spatial scale (or scales) using the wavelet transform, resizing

the image (to 768x768, subtending 11° visual angle), and then adjusting the luminance

values so that the resampled amplitude spectra matched those from the original

natural images. More specifically:

- To degrade a single spatial scale of natural image structure (factor N1), we

  permuted the coefficients associated with one of two spatial scales (i.e., levels):

  fine (S1) and coarse (S2). Note that the coefficient levels corresponding to fine

  and coarse natural image structure are dependent on the input image size and

  were determined empirically. For this, we permuted the coefficients across a

series of levels and chose the two levels corresponding to fine and coarse natural image structure by visual inspection. Note that the fine scale manipulation targeted structure in the range of 4.5 - 8.8 cycles per degree and the coarse scale manipulation targeted structure in the range of 1.3 – 2.4 cycles per degree.

- To construct noise images that shared the same basic image properties as our natural images (factor N2), we simply performed the wavelet degradation on the natural images across all spatial scales. This destroys all natural image structure, leaving a noise image with the same $1/f^\alpha$ frequency distribution as the original natural image.

- To put natural image back into a noise image (factor N3), we first degraded all the spatial scales except that of interest (i.e., all but S1 or S2). Then we degraded the remaining structure at that scale. This produced a pair of images: one noise image (all scales permuted) and another that was identical to the noise image except that one spatial scale of information still remained.

All wavelet resampling was performed using Daubechies wavelets, which are a family of orthogonal wavelets characterized by a maximal number of vanishing moments while minimising asymmetry (here we used the db6 wavelet with 6 vanishing moments). To avoid edge effects when performing the wavelet degrading, which manifest as sharp horizontal or vertical striping in the image, we did not perform the wavelet degradation over the entire image. Instead, we left an outer border (1/20th of the image size) untouched around the entire image. After the detail coefficients associated with spatial locations inside this border were permuted, the image was cropped so that only the permuted portion remained.

557

3.1.4 Retinotopic localizer

To localize cortical responses to visual images, we performed two types of phase-encoded retinotopic mapping: one to map polar angle and the other to map eccentricity representations. Briefly, the polar angle stimulus consisted of a rotating bowtie (two wedges opposite one another and meeting at fixation) and the eccentricity stimulus consisted of an expanding ring (Schira et al., 2009). The aperture contained one of three colored texture patterns (checkers, expanding and contracting spirals, or rotating sinusoidal gratings) which changed randomly every 250 ms. Participants performed a fixation color detection task at a central maker, and a fixation grid was overlaid atop the stimuli.

3.1.5. Magnetic resonance imaging data acquisition

Data were acquired on a Philips 3T Achieva X Series equipped with Quasar Dual gradients and a 32-channel head coil. Whole-brain, anatomical images were collected using a magnetisation-prepared rapid acquisition with gradient echo MPRAGE sequence with a TE of 2.8 ms, TR of 6.3 ms, flip angle of 8 degrees, FOV of 256 mm x 256 mm, a matrix size of 340 x 340, and 250 slices that were 0.75 mm thick – resulting in an isotropic voxel size of 0.75 mm.

The voxel resolution of the functional echo planar images (EPIs) collected here was 1.5 x 1.5 x 1.5 mm^3 across 31-32 oblique coronal slices covering the occipital pole. EPIs were acquired with a TR of 2 s, a TE of 25 ms, a SENSE factor of 2, a 128 x 128 matrix (ascending acquisition), and a FOV of 192 mm. For polar angle mapping 186 volumes were collected, for eccentricity mapping 174 volumes were collected, and for

582   the natural image experiment 184 volumes were collected. Before data analysis, the

583   first few volumes were discarded to account for the high T1 saturation that occurs at

584   the beginning of a scan. For both mapping protocols the first 6 volumes were

585   discarded, and for the natural image experiment the first 4 volumes were discarded.

586

587   *3.1.6 Data analysis*

588   Pre-processing of the functional data was performed using SPM8 (SPM software

589   package, Wellcome Department, London, UK; http://www.fil.ion.ucl.ac.uk/spm/). Data

590   were motion corrected using a rigid body transform and 7th degree B-spline

591   interpolation. Images were slice scan time corrected using the first image as the

592   reference slice and resliced into the space of the first image.

593

594   For retinotopic mapping, "traveling-wave" analysis procedures were conducted using

595   the   mrVista   Toolbox   (Stanford   University,   Stanford,   CA;

596   http://white.stanford.edu/software/). The cyclic retinotopic mapping data was analysed

597   using a fast Fourier transform based correlation analysis, as built in the mrLoadRet

598   software from the mrVISTA toolbox. This estimates a coherency value for each voxel

599   in the cortex as a ratio between the power at the stimulus frequency and noise. The

600   retinotopic location (both polar angle and eccentricity) for each voxel was determined

601   by the phase value at the stimulus frequency. The retinotopy data were then displayed

602   on a 3D rendered brain surface (Engel et al., 1997; Schira et al., 2009).

603

604   Volumetric segmentation of white matter was performed manually using ITK Gray

605   (Yushkevich et al., 2006). 3D surface reconstructions of the left and right hemisphere

606   were generated using mrMesh (a function within the mrVista Toolbox) by growing a 3-

607   voxel thick layer (1.5 mm isotropic voxels) above the grey/white boundary. To improve

608   data visualisation (i.e. when projecting functional data onto surfaces), these surfaces

609   were also computationally-inflated using the "smoothMesh" option in mrMesh (8

610   iterations). Note that the cortical surface models were only used for data visualization

611   and region-of-interest (ROI) definition. All analyses and statistics were performed

612   using the volumetric data.

613

614   Further analysis in the mrVista Toolbox included a general linear model (GLM) of

615   responses across early visual areas (V1, V2, V3) for each individual subject. The

616   Boynton Gamma HRF was used to model the haemodynamic response function

617   (Boynton et al., 1996). All runs were concatenated and the null grey background

618   condition was used as baseline.

619

620   <u>3.2 Results</u>

621   We first used the retinotopic mapping data to define V1, V2, and V3 ROIs in both

622   hemispheres for each individual (Fig. 10). We then extracted the GLM-derived β-

623   weights associated with each experimental condition from all voxels in each ROI. The

624   mean β-weight was then computed for each visual area, combining both hemispheres.
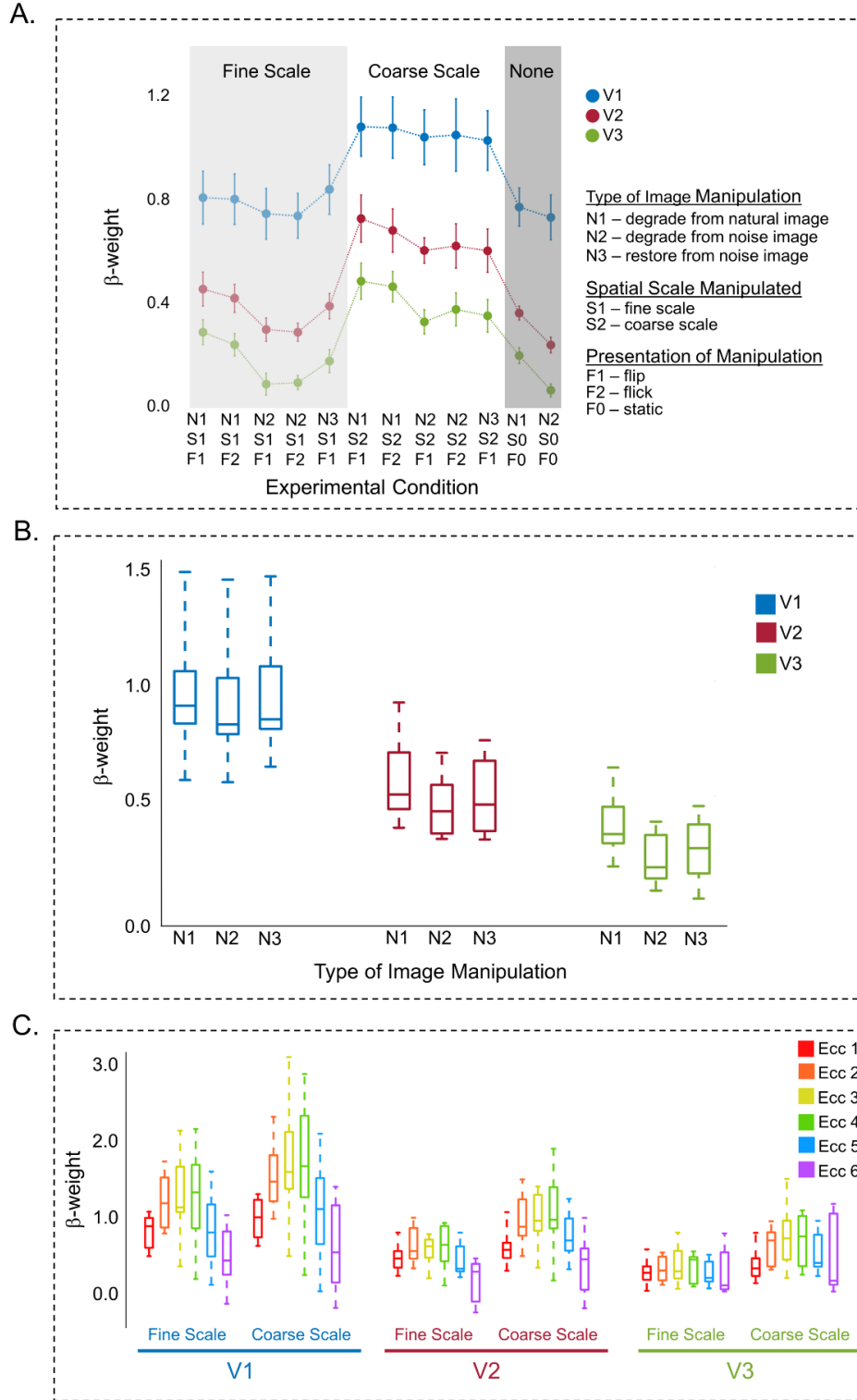
625

626   Figure 11A shows the average response in each of the visual area ROIs for each

627   condition across all subjects. Inspection of Figure 11A reveals a few salient, interesting

628   response differences across visual areas and across experimental conditions.

629   Notably, as one progresses up the visual hierarchy (V1→V2→V3), the response

630   amplitude decreases across all conditions. It also appears that, in general, the natural

631   images elicit greater activation than the noise images (N1>N2,N3). This is true not

only for the conditions involving image manipulation, but also for the no manipulation conditions (N1,S0,F0 vs. N2,S0,F0). However, the degree of difference between natural image (N1) vs. noise image (N2) conditions appears to become greater as one progresses up the hierarchy.
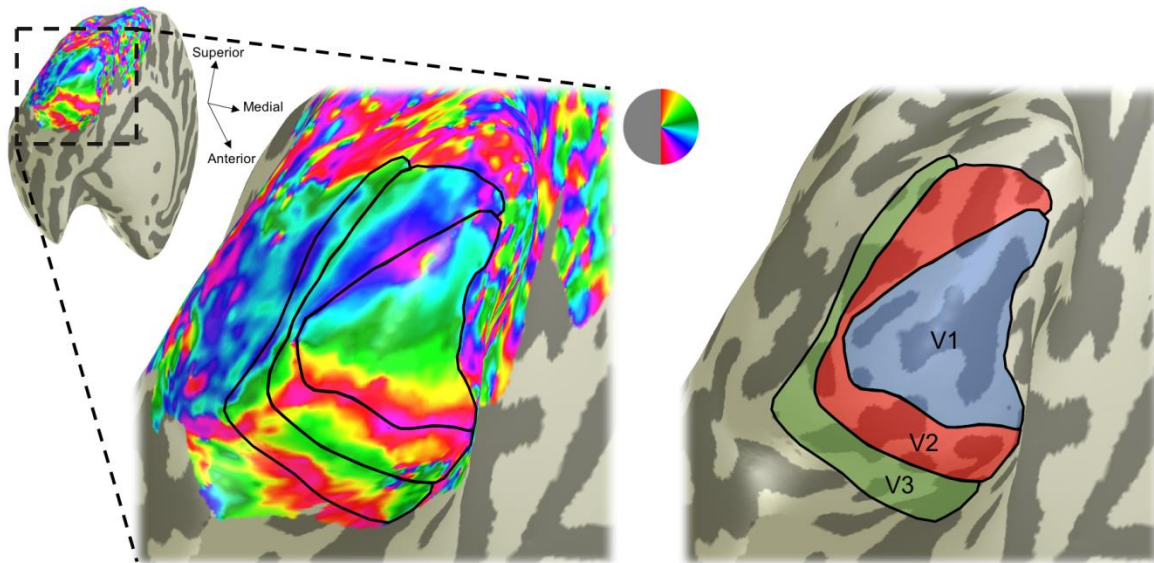
Figure 10. Defining visual area ROIs. For each individual subject, early visual cortex was partitioned into V1, V2, and V3 ROIs using polar angle retinotopic mapping data. On the far left is an inflated cortical surface model for the left hemisphere of a single subject. Next to that is a zoomed-in view of the occipital cortex showing the polar angle retinotopic map (un-thresholded). On the right is the same zoomed-in view of the occipital cortex, showing the three visual area ROIs overlaid upon the curvature pattern.

Figure 11. Activation across the early visual hierarchy for intact vs. degraded natural images. (A) Group averaged β-weights for all experimental conditions in each visual areas ROI. Error bars represent SEM across individuals. (B) β-weights in each visual area ROI for the different types of image manipulations (N1: degrade from natural image, N2: degrade from noise, N3: restore from noise), collapsed across all other factors. (C) β-weights across eccentricity for both scale conditions in each visual area, collapsed across other factors. Ecc 1 to Ecc 6 range from the fovea to the periphery ($0.06 \leq$ Ecc 1 ≤ 0.48; $0.48 \leq$ Ecc 2 ≤ 0.95; $0.95 \leq$ Ecc 3 ≤ 1.36; $1.36 \leq$ Ecc 4 ≤ 1.93; $1.93 \leq$ Ecc 5 ≤ 2.74; $2.74 \leq$ Ecc 6 ≤ 3.89 degrees). For (B) and (C), whiskers with caps show min/max, bottom and top edges of boxes indicate 25th and 75th percentile, and central line marks the median across all participants.

653

Qualitative assessment of Figure 11A appears to support the core hypothesis that higher cortical areas are more sensitive to more complex statistical features of natural scenes than V1 (i.e., cortical areas respond more strongly when natural image structure is present than when absent and this difference increases as one progresses up the hierarchy). To test this, we collapsed the data across the spatial scale (S1 and S2) and presentation (F1 and F2) factors, and removed the static, no manipulation conditions (N1,S0,F0 and N2,S0,F0; Figure 11B). We then performed a 2-way repeated measures ANOVA to investigate if the visual areas differentially responded to the different image manipulations (N1, N2, and N3). We found that a differential response was indeed present. That is, in addition to significant main effects for both visual area [F=52.3, p<0.001] and the type of image manipulation [F=11.2, p=0.0018], there was also a significant interaction effect [F=16.6, p<0.001]. Looking at Figure 11B, it appears that the interaction effect reflects an increasing effect of natural image structure on the responses as one progresses from V1 to V3. Recall that N1 is a natural image with one scale of structure degraded, N2 is essentially a noise image (all scales of structure degraded), and N3 is mostly a noise image but still has one scale of structure present. Hence, N1 has the most natural image structure, N3 the second most, and N2 has the least. In V1, there is little difference among the three conditions suggesting that V1 is only weakly influenced by the presence versus absence of the higher-order correlations that characterize natural image structure. In V2, however, the effect of image type on the average response becomes stronger and appears graded by the amount of structure present. This same differential response is further pronounced in V3.

677

678　Sensitivity to different spatial scales is known to vary as functions of both visual area

679　and eccentricity. That is, receptive field size increases up the visual hierarchy and at

680　increasingly peripheral eccentricities (Dumoulin and Wandell, 2008). We hence also

681　explored the effect of the scale condition (S) and its interaction with visual area and

682　eccentricity. For this, we first sub-divided each visual area ROI into 6 eccentricity

683　bands using the retinotopic mapping data ($0.06 \leq$ Ecc 1 ≤ 0.48; $0.48 \leq$ Ecc 2 ≤ 0.95;

684　$0.95 \leq$ Ecc 3 ≤ 1.36; $1.36 \leq$ Ecc 4 ≤ 1.93; $1.93 \leq$ Ecc 5 ≤ 2.74; $2.74 \leq$ Ecc 6 ≤ 3.89

685　degrees). We collapsed the data across the presentation (F1 and F2) and image

686　manipulation (N1, N2, and N3) factors, and removed the static, no manipulation

687　conditions (Figure 11C). We then performed a 3-way repeated measures ANOVA

688　finding a significant main effect again for visual area [$F=37.0$, $p<0.001$] as well as

689　significant main effects for eccentricity [$F=3.5$, $p=0.014$] and scale [$F=80.6$, $p<0.001$].

690　There were also significant interaction effects between visual area and eccentricity

691　[$F=3.3$, $p=0.002$] as well as between eccentricity and scale [$F=4.5$, $p=0.004$] but not

692　between visual area and scale [$F=1.3$, $p=0.303$] nor among the three [$F=0.6$, $p=0.804$].

693　Looking at Figure 11C, the main effects are clear. For visual area, we see a general

694　diminishing of the response as one progresses up the visual hierarchy (similar to the

695　effect of area seen in Figure 11B). For eccentricity, we see a the same basic inverted-

696　U pattern across eccentricity for each combination of spatial scale condition and visual

697　area except for the fine scale condition in V3 (likely driving the interaction effect). For

698　the scale condition, we see consistently greater responses to the coarse scale

699　manipulation compared to the fine scale (also clearly seen in Figure 11A), particularly

700　at intermediate eccentricities.

701

702 With respect to the scale effect, note that the process of wavestrapping a noise image

703 (N2) simply results in another noise image since no structure was originally present.

704 However, it is important to understand that the resulting noise image is still different

705 from the source noise image, and the difference is dependent on the manipulated

706 scale. Therefore, when the images are presented by flicking between or flipping

707 through the different instances, changes in the image occur at the targeted spatial

708 scale. From our results then, it appears that when the changes occur at the coarse

709 scale, a higher degree of activity is seen in visual cortex compared to when the

710 changes occur at the fine scale. The perceptual difference between the fine and

711 coarse scale resampling of noise can be seen by contrasting conditions N2,S1,F1 vs.

712 N2,S2,F1 or N2,S1,F2 vs. N2,S2,F2 in Supplementary Movie 2.

713

714 Note that the primary motivation for 'flipping' or 'flicking' across multiple instances

715 within a block was to make the stimuli "dynamic" and hence more salient to the visual

716 system compared to using a static image across the block duration. The choice of

717 flipping versus flicking was selected to probe the role of prior context on visual

718 responses – i.e. whether a statistical violation (the wavelet-degraded scale) would

719 have a greater cortical salience when introduced in and out of a preserved scene (F1),

720 or whether the violation would accrue a stronger response when continually presented

721 (F2). Whereas the dynamic conditions did elicit greater responses than their

722 corresponding static conditions (Figure 11A), we did not find any main effect of the

723 presentation factor (F1 vs. F2) [F=0.2, p=0.681] nor an interaction with visual area

724 [F=3.4, p=0.067] when conducting a 2-way repeated measures ANOVA.

725

726 **4. Discussion**

727 Sensory and cognitive neuroscience has traditionally employed simple, abstract, and

728 narrowband stimuli to examine cortical response properties. These stimuli have

729 served the field well, offering a way to tightly control variables of interest and leading

730 to an extensive characterization of the response of single neurons and populations of

731 neurons to basic image properties such as luminance, contrast, orientation, and

732 spatial frequency. Despite this, these stimuli lack ecological validity as they rarely

733 come close to approximating the types of stimuli encountered in typical sensory

734 experiences outside of experimental conditions. Pertinently, there is mounting

735 evidence suggesting that the cortex may be more strongly 'tuned' to the statistical

736 properties of naturalistic stimuli (for review, see Sonkusare et al., (2019)). For

737 example, a recent study (Isherwood et al., 2017) using broadband noise stimuli

738 observed that stimuli with $1/f^{\alpha}$ spectra close to that of natural scenes (i.e., $\alpha = 1.25$,

739 Figure 1C) elicited stronger BOLD responses than stimuli with $1/f^{\alpha}$ spectra outside of

740 the natural range (i.e., $\alpha = 0.25$ or $\alpha = 2.25$). Interestingly, this apparent tuning of the

741 cortex to the spectra of natural stimuli is mirrored by visual sensitivity and preference

742 at the behavioral level. Discrimination sensitivity, detection sensitivity, as well as

743 aesthetic preference are highest for noise stimuli with natural $1/f^{\alpha}$ spectra and lowest

744 for unnatural $1/f^{\alpha}$ spectra (Spehar and Taylor, 2013; Spehar et al., 2015). This

745 supports the notion that the visual system is tuned to the statistical properties of natural

746 scenes. Findings such as these highlight the importance of using more complex,

747 naturalistic stimuli in neuroscientific pursuits.

748

749 The benefit of complementing studies using traditional, abstract stimuli with those that

750 use more ecological stimuli is clear. The use of naturalistic stimuli, however, is still

751 relatively nascent, and as such, considerable challenges remain. One such issue is

752 determining how to manipulate these naturalistic stimuli with sufficient control and

753 rigor. Seminal early work disrupted the temporal narrative by sharp block shuffling of

754 movie segments in the time domain to unveil large-scale temporal hierarchies in the

755 cortex (Hasson et al., 2008). The wavelet approach outlined in the present manuscript

756 offers an alternative, more nuanced opportunity in this direction to turn the focus on

757 hierarchies in the visual system. Our work demonstrates that it is possible to

758 parametrically and subtly manipulate the complex statistical properties of natural

759 scenes with a high degree of control and flexibility – and that the visual system is

760 sensitive to these subtle manipulations.

761

762 There are a wide range of ways that wavelets can be used to manipulate stimuli to

763 probe functional effects of natural scene statistics in the visual hierarchy, some of

764 which were described in Part 1. The neuroimaging study here (Part 2) makes use of

765 one of these, demonstrating some of the practical considerations of performing an

766 fMRI experiment using wavelet-degraded stimuli. In doing so, we found evidence in

767 support of our main hypothesis (that higher hierarchical regions in visual cortex are

768 more sensitive to natural scene statistic). These results are convergent with other

769 recent research, using substantially different visual stimuli, showing that sensitivity to

770 the distinctive higher-order correlations of natural scenes begins to arise in visual area

771 V2. For example, Freeman et al., (2013) found that generated, naturalistic texture

772 stimuli (with higher-order correlations) differentially modulated cortical responses in

773 V2 but not V1 compared to spectrally matched noise (without the higher-order

774 correlations). Notably, comparable results were found by the authors using both fMRI

775 in humans and neural recordings in macaque. Yu et al. (2015) similarly showed that

776 many neurons in macaque V2 (but few in V1) are sensitive to higher-order properties

777 of natural scenes. Rather than degrading natural images as done in the present study

778 or constructing stimuli that mimic naturalistic textures (Freeman et al., 2013), Yu et al.

779 used binary textures that were highly unnatural, but isolated specific multipoint

780 correlations characteristic of natural images (i.e., the statistics of the combinations of

781 luminance values that appear in several points of a natural image) (Hermundstad et

782 al., 2014; Tkacik et al., 2010). Note that the uniform textures generated by Freeman

783 et al. (2013) appears more "natural" than the binary textures (Yu et al., 2015), although

784 both can be easily visually disambiguated from an actual natural image as they lack

785 the contextual information and complex variability present in natural scenes. It is clear

786 then, that although selectivity to higher-order correlations in natural images begin to

787 arise in V2, future work is required to determine where along the hierarchy further

788 selectivity to additional natural image structure emerges.

789

790 The human visual system is composed of many functionally distinct cortical visual

791 areas (Grill-Spector and Malach, 2004; Zeki et al., 1991). Sensory-driven responses

792 tend to decrease as one progress up the visual hierarchy, and as such, our finding

793 that responses to all of our stimuli decrease as one progresses up the visual hierarchy

794 is unsurprising. Notably, however, we also found that the higher cortical areas appear

795 to be more sensitive to the complex visual features – that is, the decrease in responses

796 up the visual stream was more pronounced for wavelet resampled stimuli. The present

797 application to fMRI data thus suggests that the higher order structure being degraded

798 by the wavelet technique is directly related to the complex features that the higher

799 visual areas encode. That is, cells along the visual hierarchy become increasingly

800 sensitive to the conditional dependences among multiple neurons in lower hierarchical

801 levels, mirroring the complex conditional dependences in unaltered natural scenes.

802   Presumably, this effect would be stronger in even higher-order areas; however, our

803   data are insufficient to test this. Due to the size and orientation of our fMRI acquisition

804   slab, we only have partial coverage of hV4 for most participants. In addition, time

805   constraints restricted the number of runs of each retinotopic mapping stimulus –

806   limiting the data quality and thus our ability to confidently demarcate higher-order

807   dorsal and lateral areas. Future studies could be designed to circumvent this issue,

808   for example by having a separate scan session dedicated to the collection of a

809   comprehensive, high-quality retinotopic mapping dataset.

810

811   One powerful aspect of the wavelet-based approach outlined here is the ability to

812   target structure at specific spatial scales. As mentioned, receptive field size and hence

813   spatial frequency sensitivity is known to vary both across visual areas as well as

814   across eccentricities within a visual area (Dumoulin and Wandell, 2008; Yoshor et al.,

815   2007). By combining the experiment with fMRI-based estimates of population

816   receptive field sizes (Dumoulin and Wandell, 2008; Zeidman et al., 2018), future

817   studies will be able to take a more detailed look at the relationship between cortical

818   activity related to specific scales of natural image structure and the underlying

819   receptive field sizes. Our preliminary results suggest that manipulations to coarse

820   scales elicit stronger results across the visual cortex than manipulations to the small

821   scales. Interestingly, this is found when wavestrapping the noise images (N2) as well

822   as those with structure present (N1). Although the mean perturbation across the

823   images and realizations do not show a scale-specific effect, the variability is higher at

824   coarser scales (see Fig. 11). Hence the greater responses to coarse scale

825   manipulations (S2) compared to the fine scale manipulations (S1) may either reflect

826   stronger neuronal sensitivity to coarse scale information or encoding of the trial-to-trial

827 variability. In studying the effect of scale, it will also be important to test across the full

828 range of spatial scales, rather than only two as done in the present study. Full-range,

829 parametric studies are necessary to reveal any important non-monotonicity that might

830 be present in the response properties (Rainer et al., 2001).

831

832 Although participants in our experiment attended to a fixation task while passively

833 viewing raw and altered static natural images presented in successive transitions, it is

834 important to note that perception in the wild is embedded in a broader action-

835 perception cycle (Fuster, 2002). It thus makes sense to not only use wavelet

836 resampling to degrade the spatial and temporal statistics, but to do so while

837 participants freely view movies (i.e., with unrestricted eye movements). As reviewed

838 above, wavelet resampling is directly applicable to dynamic, spatio-temporal stimuli

839 (S2, Sup Movie 1) – and there exists several different ways of achieving this:

840 preserving, destroying, or manipulating the complex temporal statistics embedded in

841 dynamic natural scenes. Block resampling is one variant of this broader class,

842 preserving the temporal structure within blocks but degrading the temporal spectra –

843 precisely and only at the time-scale of the block.

844

845 As a final consideration, image manipulations of higher order statistics could be made

846 at the time of saccades, during fixational eye movements, or during scene transitions

847 – introducing subtle stimulus errors into the active stream of visual perception, while

848 avoiding low-level changes in luminance, contrast, or spectra. This inclusion of

849 parametric prediction errors would allow novel probes of the predictive coding

850 principles of visual function (Edwards et al., 2017; Friston, 2005; Vetter et al., 2012).

851 Other recent work has used wavelet resampling to construct dynamic stimuli from a

static natural scene by cyclically permuting the wavelet scales, hence tuning a static scene in and out of its (preserved) noise context (Koenig-Robert and VanRullen, 2013; Koenig-Robert et al., 2015). This approach allows cyclic presentation of both expected and surprising semantic content (of the natural scene) while keeping the spectral properties of the stimulus constant (unlike a traditional event related paradigm), thus probing cortical hierarchies for their role in predictive coding and error responses (Gordon et al., 2019a; Gordon et al., 2017; Gordon et al., 2019b).

References

866

867 Allison, T., Puce, A., McCarthy, G., 2000. Social perception from visual cues: role of

868   the STS region. Trends Cogn Sci 4, 267-278.

869 Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J., 1996. Linear systems analysis

870   of functional magnetic resonance imaging in human V1. J Neurosci 16, 4207-

871   4221.

872 Breakspear, M., Brammer, M., Robinson, P.A., 2003. Construction of multivariate

873   surrogate sets from nonlinear data using the wavelet transform. Physica D 182,

874   1-22.

875 Breakspear, M., Brammer, M.J., Bullmore, E.T., Das, P., Williams, L.M., 2004.

876   Spatiotemporal wavelet resampling for functional neuroimaging data. Hum

877   Brain Mapp 23, 1-25.

878 Bullmore, E., Fadili, J., Breakspear, M., Salvador, R., Suckling, J., Brammer, M., 2003.

879   Wavelets and statistical analysis of functional magnetic resonance images of

880   the human brain. Stat Methods Med Res 12, 375-399.

881 Bullmore, E., Long, C., Suckling, J., Fadili, J., Calvert, G., Zelaya, F., Carpenter, T.A.,

882   Brammer, M., 2001. Colored noise and computational inference in

883   neurophysiological (fMRI) time series analysis: resampling methods in time and

884   wavelet domains. Hum Brain Mapp 12, 61-78.

885 Burton, G.J., Moorhead, I.R., 1987. Color and spatial structure in natural scenes. Appl

886   Opt 26, 157-170.

887 Cocchi, L., Gollo, L.L., Zalesky, A., Breakspear, M., 2017. Criticality in the brain: A

888   synthesis of neurobiology, models and cognition. Prog Neurobiol 158, 132-152.

889 Dumoulin, S.O., Wandell, B.A., 2008. Population receptive field estimates in human

890   visual cortex. Neuroimage 39, 647-660.

891   Edwards, G., Vetter, P., McGruer, F., Petro, L.S., Muckli, L., 2017. Predictive feedback
892   to V1 dynamically updates with sensory input. Sci Rep 7, 16538.

893   Einhauser, W., Konig, P., 2003. Does luminance-contrast contribute to a saliency map
894   for overt visual attention? Eur J Neurosci 17, 1089-1097.

895   Engel, S.A., Glover, G.H., Wandell, B.A., 1997. Retinotopic organization in human
896   visual cortex and the spatial precision of functional MRI. Cereb Cortex 7, 181-
897   192.

898   Field, D.J., 1987. Relations between the statistics of natural images and the response
899   properties of cortical cells. J Opt Soc Am A 4, 2379-2394.

900   Field, D.J., 1994. What is the goal of sensory coding? Neural Computation 6, 559-601.

901   Field, D.J., 1999. Wavelets, vision and the statistics of natural scenes. Roy Soc of
902   London Phil Tr A 357, 2527.

903   Frazor, R.A., Geisler, W.S., 2006. Local luminance and contrast in natural images.
904   Vision Res 46, 1585-1598.

905   Freeman, J., Ziemba, C.M., Heeger, D.J., Simoncelli, E.P., Movshon, J.A., 2013. A
906   functional and perceptual signature of the second visual area in primates. Nat
907   Neurosci 16, 974-981.

908   Friston, K., 2005. A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci
909   360, 815-836.

910   Fuster, J.M., 2002. Physiology of executive functions: The perception-action cycle. In:
911   Stuss, D.T., Knight, R.T. (Eds.), Principles of frontal lobe function. Oxford
912   University Press, New York, NY, US pp. 96-108.

913   Gordon, N., Hohwy, J., Davidson, M.J., van Boxtel, J.J.A., Tsuchiya, N., 2019a. From
914   intermodulation components to visual perception and cognition-a review.
915   Neuroimage 199, 480-494.

916 Gordon, N., Koenig-Robert, R., Tsuchiya, N., van Boxtel, J.J., Hohwy, J., 2017. Neural
917   markers of predictive coding under perceptual uncertainty revealed with
918   Hierarchical Frequency Tagging. Elife 6.

919 Gordon, N., Tsuchiya, N., Koenig-Robert, R., Hohwy, J., 2019b. Expectation and
920   attention increase the integration of top-down and bottom-up signals in
921   perception through different pathways. PLoS Biol 17, e3000233.

922 Graham, D., Schwarz, B., Chatterjee, A., Leder, H., 2016. Preference for luminance
923   histogram regularities in natural scenes. Vision Res 120, 11-21.

924 Graps, A., 1995. An introduction to wavelets. IEEE Comput Sci Eng 2, 50-61.

925 Grill-Spector, K., Malach, R., 2004. The human visual cortex. Annu Rev Neurosci 27,
926   649-677.

927 Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject
928   synchronization of cortical activity during natural vision. Science 303, 1634-
929   1640.

930 Hasson, U., Yang, E., Vallines, I., Heeger, D.J., Rubin, N., 2008. A hierarchy of
931   temporal receptive windows in human cortex. J Neurosci 28, 2539-2550.

932 Hermundstad, A.M., Briguglio, J.J., Conte, M.M., Victor, J.D., Balasubramanian, V.,
933   Tkacik, G., 2014. Variance predicts salience in central sensory processing. Elife
934   3.

935 Hubel, D.H., Wiesel, T.N., 1959. Receptive fields of single neurones in the cat's striate
936   cortex. J Physiol 148, 574-591.

937 Hubel, D.H., Wiesel, T.N., 1968. Receptive fields and functional architecture of
938   monkey striate cortex. J Physiol 195, 215-243.

939　Isherwood, Z.J., Schira, M.M., Spehar, B., 2017. The tuning of human visual cortex to

940　　　variations in the 1/f(alpha) amplitude spectra and fractal properties of synthetic

941　　　noise images. Neuroimage 146, 642-657.

942　Karklin, Y., Lewicki, M.S., 2009. Emergence of complex cell properties by learning to

943　　　generalize in natural scenes. Nature 457, 83-86.

944　Koenig-Robert, R., VanRullen, R., 2013. SWIFT: a novel method to track the neural

945　　　correlates of recognition. Neuroimage 81, 273-282.

946　Koenig-Robert, R., VanRullen, R., Tsuchiya, N., 2015. Semantic Wavelet-Induced

947　　　Frequency-Tagging (SWIFT) Periodically Activates Category Selective Areas

948　　　While Steadily Activating Early Visual Areas. PLoS One 10, e0144858.

949　Munn, B., Gong, P., 2018. Critical Dynamics of Natural Time-Varying Images. Phys

950　　　Rev Lett 121, 058101.

951　Olshausen, B.A., Field, D.J., 1996. Emergence of simple-cell receptive field properties

952　　　by learning a sparse code for natural images. Nature 381, 607-609.

953　Puckett, A.M., DeYoe, E.A., 2015. The attentional field revealed by single-voxel

954　　　modeling of fMRI time courses. J Neurosci 35, 5030-5042.

955　Rainer, G., Augath, M., Trinath, T., Logothetis, N.K., 2001. Nonmonotonic noise tuning

956　　　of BOLD fMRI signal to natural images in the visual cortex of the anesthetized

957　　　monkey. Curr Biol 11, 846-854.

958　Roberts, J.A., Wallis, G., Breakspear, M., 2013. Fixational eye movements during

959　　　viewing of dynamic natural scenes. Front Psychol 4, 797.

960　Ruderman, D.L., Bialek, W., 1994. Statistics of natural images: Scaling in the woods.

961　　　Phys Rev Lett 73, 814-817.

962　Saremi, S., Sejnowski, T.J., 2013. Hierarchical model of natural images and the origin

963　　　of scale invariance. Proc Natl Acad Sci U S A 110, 3071-3076.

964 Schira, M.M., Tyler, C.W., Breakspear, M., Spehar, B., 2009. The foveal confluence

965   in human visual cortex. J Neurosci 29, 9050-9058.

966 Schira, M.M., Wade, A.R., Tyler, C.W., 2007. Two-dimensional mapping of the central

967   and parafoveal visual field to human visual cortex. J Neurophysiol 97, 4284-

968   4295.

969 Schultz, J., Pilz, K.S., 2009. Natural facial motion enhances cortical responses to

970   faces. Exp Brain Res 194, 465-475.

971 Sekuler, A.B., Bennett, P.J., 2001. Visual neuroscience: Resonating to natural images.

972   Curr Biol 11, R733-736.

973 Sonkusare, S., Breakspear, M., Guo, C., 2019. Naturalistic Stimuli in Neuroscience:

974   Critically Acclaimed. Trends Cogn Sci 23, 699-714.

975 Spehar, B., Taylor, R.P., 2013. Fractals in art and nature: why do we like them? ,

976   Human Vision and Electronic Imaging XVIII. International Society for Optics and

977   Photonics, p. 865118.

978 Spehar, B., Wong, S., van de Klundert, S., Lui, J., Clifford, C.W., Taylor, R.P., 2015.

979   Beauty and the beholder: the role of visual sensitivity in visual preference. Front

980   Hum Neurosci 9, 514.

981 Tkacik, G., Prentice, J.S., Victor, J.D., Balasubramanian, V., 2010. Local statistics in

982   natural scenes predict the saliency of synthetic textures. Proc Natl Acad Sci U

983   S A 107, 18149-18154.

984 Tolhurst, D.J., Tadmor, Y., Chao, T., 1992. Amplitude spectra of natural images.

985   Ophthalmic Physiol Opt 12, 229-232.

986 Treisman, A.M., Gelade, G., 1980. A feature-integration theory of attention. Cogn

987   Psychol 12, 97-136.

988 van der Schaaf, A., van Hateren, J.H., 1996. Modelling the power spectra of natural
989   images: statistics and information. Vision Res 36, 2759-2770.

990 Vetter, P., Edwards, G., Muckli, L., 2012. Transfer of predictive signals across
991   saccades. Front Psychol 3, 176.

992 Yoshor, D., Bosking, W.H., Ghose, G.M., Maunsell, J.H., 2007. Receptive fields in
993   human visual cortex mapped with surface electrodes. Cereb Cortex 17, 2293-
994   2302.

995 Yu, Y., Schmid, A.M., Victor, J.D., 2015. Visual processing of informative multipoint
996   correlations arises primarily in V2. Elife 4, e06604.

997 Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G.,
998   2006. User-guided 3D active contour segmentation of anatomical structures:
999   significantly improved efficiency and reliability. Neuroimage 31, 1116-1128.

1000 Zeidman, P., Silson, E.H., Schwarzkopf, D.S., Baker, C.I., Penny, W., 2018. Bayesian
1001   population receptive field modelling. Neuroimage 180, 173-187.

1002 Zeki, S., Watson, J.D., Lueck, C.J., Friston, K.J., Kennard, C., Frackowiak, R.S., 1991.
1003   A direct demonstration of functional specialization in human visual cortex. J
1004   Neurosci 11, 641-649.

Supplementary Material

S1: Wavelet resampling of natural images

<u>S1.1 Multiresolution wavelet decomposition</u>

1008   Wavelet resampling has been described before (Breakspear et al., 2003) but is given

1009   here for completeness. Suppose we have a grey-scaled image sampled on a two-

1010   dimensional grid over a finite domain, yielding a pixel-wise matrix of non-negative

1011   amplitude intensities, $\xi(x, y)$. The technique commences with a multi-resolution

1012   decomposition of the image ξ by the discrete wavelet transform (DWT). This is a

1013   representation of the image across a hierarchy of spatial and/or temporal scales. At

1014   each scale j, the image is decomposed into two orthogonal components: the detail

1015   coefficients $d_{j,k}$, which contain information about the fluctuations in signal intensity at

1016   that scale, and the approximation coefficients $a_{j,k}$, which represent the residual of the

1017   signal after those and all smaller details have been removed. The index k is a

1018   translation matrix which codes the position of the coefficients in space. The original

1019   signal can be recovered without loss by linearly adding the approximation of the signal

1020   at any scale together with the details at that and all smaller scales.

1021   Wavelets are families of basis functions that permit such a decomposition. A family of

1022   wavelet functions $\{\varphi_{j,k}\}$ is generated through dilation (by scale factor j) and translation

1023   (by position factor k) of a single "mother" wavelet function φ. Uniquely associated with

1024   each mother wavelet function is a family of scaling functions $\{\phi_{j,k}\}$ generated by

1025   dilation and translation of a single "father" scaling function ϕ. Wavelet and scaling

1026   functions have a finite support and code image intensity at a specific scale and local

1027   position. Convolution of the signal with the wavelet functions $\varphi_{j,k}$ produces the detail

1028 coefficients $d_{j,k}$. Convolution with the scaling functions $\phi_{j,k}$ produces the approximation

1029 coefficients $d_{j,k}$. Hence, a multi-scale decomposition of an image ξ at scale J is given

1030 by

1031
$$\xi = \sum_k a_{J,k}\phi_{J,k} + \sum_{j\leq J}\sum_k d_{j,k}\varphi_{j,k}.$$

1032 For the discretely sampled image $\xi(x,y)$, the coefficients vanish outside of a closed

1033 interval, and hence the number of terms in the sum is finite. We use the notation that

1034 $j = 1$ is the smallest scale (determined by the sampling frequency). The DWT is a

1035 dyadic decomposition in the sense that the wavelet functions at each scale are dilated

1036 by a factor of two between each level (from fine to coarse) and the number of

1037 coefficients is halved in each direction. For a one-dimensional signal $\xi(x)$ of length S,

1038 the number of detail coefficients N_j at scale j is,

1039
$$N_j = S/2^j + l,$$

1040 where l is a small integer that allows the edges of the signal to be covered and

1041 depends on the support width of the wavelet functions. For the two-dimensional image

1042 $\xi(x,y)$, the detail coefficients are further decomposed into horizontal $d_{j,k}^H$, vertical $d_{j,k}^V$,

1043 and diagonal $d_{j,k}^D$ components. There are hence $3 \times (N_j)^2$ coefficients at each scale of

1044 a decomposition of a two-dimensional data set.

1045 S1.2 Two-Dimensional Wavestrapping

1046 For certain classes of random processes, the wavelet transform whitens (decorrelates)

1047 the data (see Bullmore et al., 2001; Bullmore et al., 2003). That is, correlations

1048 between nearby detail coefficients $d_{j,k}^D$ and $d_{j,k\pm1}^D$ are much weaker (and possibly

1049 uncorrelated) than correlations between neighbouring image pixels $\xi(x, y)$ and

1050 $\xi(x \pm 1, y \pm 1)$. As a result, the wavelet coefficients can be considered "exchangeable"

1051 in the sense that they can be permuted amongst themselves without destroying

1052 correlations within the reconstructed data. This property of exchangeability of wavelet

1053 coefficients is a key criterion for validity of wave-strapping schemes. In short, wave-

1054 strapping in its simplest form proceeds by wavelet transform of a spatial image,

1055 followed by random permutation of detail coefficients within each level of the

1056 decomposition; and then inverse wavelet transform of the permuted coefficients. That

1057 is, within each scale j, we take a random permutation of the set of translation matrices,

1058 $k' = \mathrm{randperm}(k)$. Then we reconstruct a wavelet degraded signal,

1059
$$\xi' = \sum_k a_{J,k} \phi_{J,k} + \sum_{j \leq J} \sum_{k'} d_{j,k'} \varphi_{j,k'}$$

1060 Note that the approximation coefficients are not decorrelated and are not permuted.

1061 Scale specific permutations are constructed by choosing one of more specific scales

1062 j' and only permuting detail coefficients within that scale,

1063
$$\xi = \sum_k a_{J,k} \phi_{J,k} + \sum_{j \neq j'} \sum_k d_{j,k} \varphi_{j,k} + \sum_{k'} d_{j',k'} \varphi_{j',k'}.$$

1064 Note that the identical permutation must be applied to the vertical, horizontal and

1065 diagonal coefficients within each scale.

1066 A color image is composed of three color palettes, $\xi = \xi^y + \xi^m + \xi^c$. To avoid color

1067 mixing, the identical permutation must be applied to each of the three colors.

1068 Numerically this can be achieved simply be resetting the random seed to the same

1069 value at the start of each separate resampling.

1070 Partial resampling is achieved by randomly selecting a proportional subset of

1071 coefficients at each scale and permuting exclusively amongst that set.

1072 For images with very strong correlations, the wavelet transform may not completely

1073 decorrelate the data. If this occurs then the reconstructed data will be whiter – i.e.,

1074 correlations amongst neighbouring pixels in the reconstructed images ξ' will be weaker

1075 than those in the original image ξ and the corresponding spectra will be flatter. Various

1076 schemes exist to address this problem (Breakspear et al., 2003). However, we did not

1077 encounter whitening with our $1/f^2$ images and used simple permutation.

1078 Because the detail coefficients are uncorrelated, they can be re-arranged according

1079 to any scheme without disrupting the correlations and spectra of the reconstructed

1080 image. Wavelet resampling classically proceeds with a random permutation,

1081 destroying higher order correlations within and between scales. "Freezing" an image

1082 in the wavelet domain proceeds by defining a single, unambiguous ordering of the

1083 coefficients that decreases the conditional entropy within and between scales.

1084 Following thermodynamic principles, we imposed an anti-magnetic scheme. This is

1085 achieved by rank ordering the coefficients at each scale from most positive to most

1086 negative. The first half of this re-ranked list are then assigned to the odd pixels, in

1087 descending order. The second half of the list are then assigned to the even pixels in

1088 ascending order. Hence neighbouring pixels are opposite in sign and the largest in

1089 magnitude are directly adjacent. This yields a clear and well-defined relationship

1090 between all coefficients at all scales.

1091 The code for these schemes are available at the authors' github repository.

S2: Examples of wavelet application and experimental stimuli

Sup Movie 1. Movie clip with spatial structure degraded using the wavelet technique. Note that the temporal structure was fully preserved by using the same random seed when degrading the structure in each frame (and each color channel within each frame).

Sup Table 1. Main experimental factors and conditions for the partial 3x2x2 within-subjects factorial design. The independent variables were type of image manipulation (N1: degrade from natural image, N2: degrade from noise, N3: restore from noise), spatial scale manipulated (S1: fine and S2: coarse), and presentation of manipulation (F1: flip vs. F2: flick).

Experimental factor			Experimental Condition
Type of image manipulation	Spatial scale manipulated	Presentation of manipulation	
N1	S1	F1	N1,S1,F1
		F2	N1,S1,F2
	S2	F1	N1,S2,F1
		F2	N1,S2,F2
N2	S1	F1	N2,S1,F1
		F2	N2,S1,F3
	S2	F1	N2,S2,F1
		F2	N2,S2,F2
N3	S1	F1	N3,S1,F1
	S2	F1	N3,S2,F1

*Note that two static image block types were also included: a natural image (N1, S0, F0) and a noise image (N2, S0, F0) along with an isoluminant grey baseline.

53

Sup Movie 2. Example of each stimulus block type used during the fMRI experiment. Note: the condition labels (bottom) were not present during the experiment.

1109

Sup Figure 1. Thumbnails showing the intact greyscale natural images used during the experiment.
See Figure S2 for the second half of the stimulus set.

1110
1111
1112

1113

1114 **Sup Figure 2.** Thumbnails showing the intact greyscale natural images used during the experiment.
1115 See Figure S1 for the first half of the stimulus set.

S3: Hierarchical model of natural images

An alternative multi-level decomposition of a grey-scaled matrix of image intensities was recently given by Saremi and Sejnowski (2013),

$$\xi = \sum_{\lambda=1}^{L} b^{L-\lambda} B^{\lambda},$$

where the integer $b > 1$ is the base of the representation, B^{λ} are non-negative matrices with values $0 < B^{\lambda} < b - 1$, and *L* is the length of the representation. B^{λ} are found iteratively, starting from $\lambda = 1$. Supplementary Figure 3 shows an example of this representation acting on a natural grey-scale image. The mid-level transition from completely ordered levels (1-3), through a partly ordered, critical phase (levels 4-6) to the disordered phase (levels 7-8) are evident.
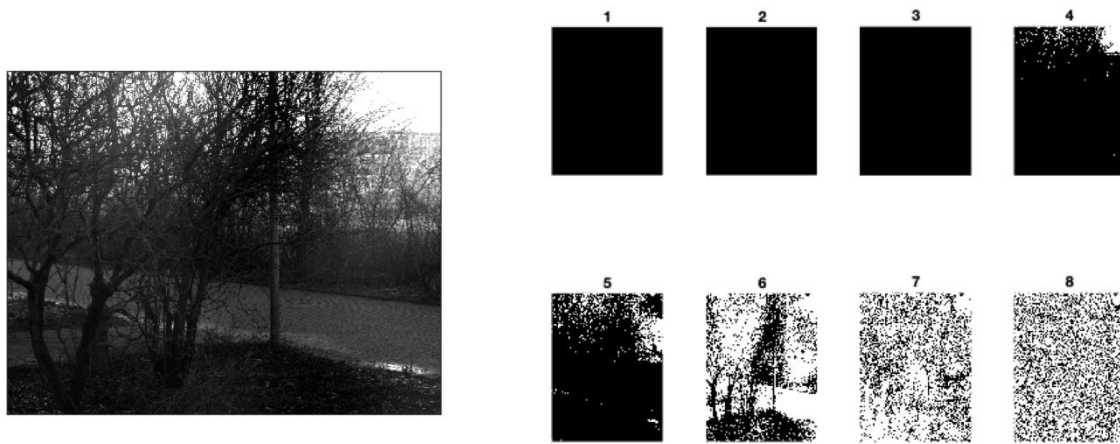
Sup Figure 3: (A) Natural image from the van Hateren image database (http://bethgelab.org/datasets/vanhateren/). (B) The 8 level decomposition showing a transition for ordered levels (1-3) to a disordered one (7-8) through an intermediate critical phase.

There are some similarities to a DWT in the sense that there is an iterative operation up to a length *L* (here) or *J* (with the DWT). However, unlike a DWT, the levels of the representation have the same dimensions as the original image ξ and are not dyadic

or downsized, even if *b*=2. The operation works only on the intensity values of the image and does not depend on the local spatial contrast. For this reason, the multi-level decomposition of a wavelet resampled image (or even a completely permuted raw image) are the same as the original image, as long as the amplitude distribution is preserved (Sup Figure 4).
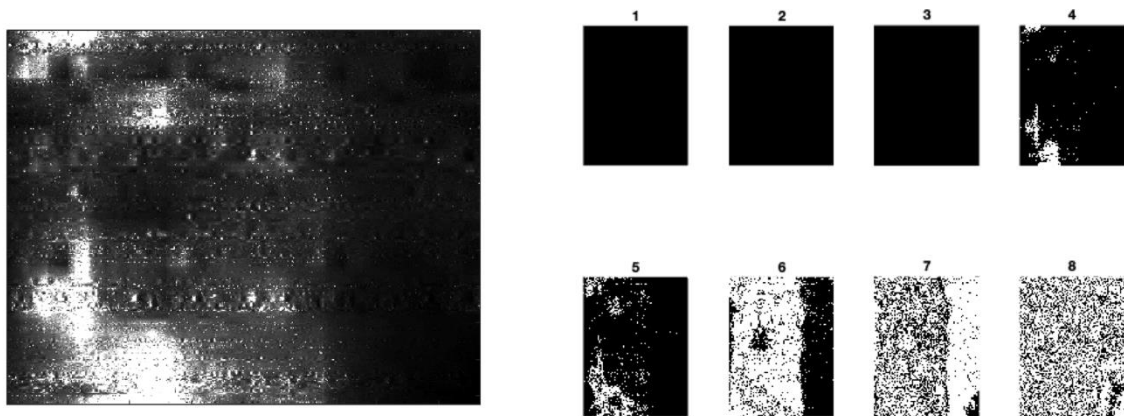
Sup Figure 4: (A) Wavelet-resampled derivative from Sup Fig 3 and (B) its 8 level decomposition.

Movie 1

Click here to access/download
Video
Movie1.avi

Click here to access/download
Video
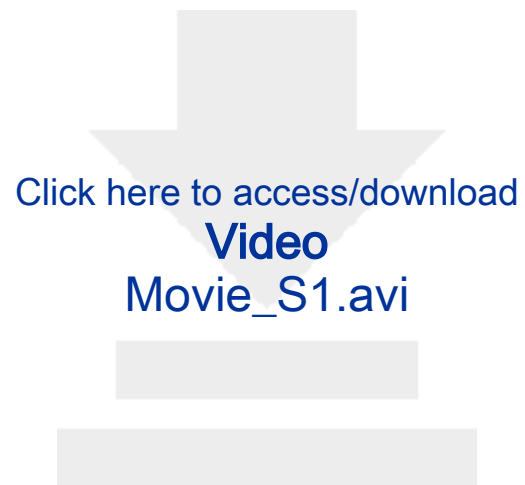Movie2.avi

Click here to access/download
Video
Movie3.avi

Click here to access/download
Video
Movie4.avi

Click here to access/download
Video
Movie5.avi

Click here to access/download
Video
Movie_S1.avi

Movie S2

Click here to access/download
Video
Movie_S2.mp4

The code used to manipulate the natural images will be made freely available by time of publication via GitHub. The original consent provided by the fMRI participants only permits data sharing among scientific collaborators of the named investigators. Please contact the corresponding author regarding data access.

The code used to manipulate the natural images will be made freely available by time of publication via GitHub. The original consent provided by the fMRI participants only permits data sharing among scientific collaborators of the named investigators. Please contact the corresponding author regarding data access.

Alexander Puckett: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization

Mark Schira: Conceptualization, Methodology, Investigation, Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition

Zoey Isherwood: Methodology, Software, Formal analysis, Investigation, Writing - Review & Editing, Visualization

Jonathan Victor: Conceptualization, Methodology, Writing - Review & Editing

James Roberts: Software, Methodology, Writing - Review & Editing, Visualization

Michael Breakspear: Conceptualization, Methodology, Software, Formal Analysis, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision