Origins of structure in perception

Sarah Schwettmann
Vision in Art and Neuroscience
2023
Disappearance of structure in perception

STABILIZED IMAGES typically fade as in the illustrations on this and the following two pages. The parts of a profile_drawn that stay visible are invariably specific features or groups of features, such as the front of the face or the top of the head.

MEANINGLESS CURVILIGUES first come and go in random sequence. But after a while, small groups of curvatures organize in recognizable patterns start to behave as units. This suggests that they have themselves become meaningful perceptual elements.

MONOGRAM formed of the letters H and B also seems to illustrate the importance of elements that are meaningful because of past experience. When the monogram breaks up it is the recognizable letters and numbers within it that come successively into view.

WORDS containing other words behave in much the same manner as the monogram. Here, for example, the subject sees new words made up of letters and parts of letters in the original. He is far less likely to report seeing meaningless groups of letters such as EER.

© 1961 SCIENTIFIC AMERICAN, INC

Stabilized Images on the Retina

When the involuntary movements of an image across the retina are prevented, the image fades and reappears in a manner that provides new information on two major theories of perception

by Roy M. Pritchard
How do the eyes receive information about the world?

425 BCE: Intromission theories

- Objects emit copies of themselves (eidola or simulacra) that enter the eyes
- Advanced by Epicurus, Lucretius, Demokritos, among others
- How could objects send copies to multiple observers? How did the copies not interfere with one another? How could copies of large objects fit into the eye?

~400 BCE: Extramission theories

- Empedocles, Plato, Euclid
- Empedocles: all things are composed of 4 elements: fire, earth, air, water. The eye contains the 4 elements and fire is responsible for creating rays that emanate from the eye and touch the world (this fire is is why cats’ eyes glow!)
- Later, Aristotle:
  - stars too far away for rays from the eyes to reach them!
  - some objects are light sources and others reflect rays that reach the eyes
  - water is the element of perception as vision needed a transparent element! (Aristotle, On Sense and the Sensible, 350 BCE)
Light travels along straight lines – coming from where?

325 BCE: Euclid, a first mathematical theory of vision (in *Optics*)

- Rays emitted *by the eye* traveled along a straight line and formed a cone reaching the scene
- Work set basis for **perspective**

From Euclid's “Optics” (325 BC)
Light travels along straight lines – coming from where?

325 BCE: Euclid, a first mathematical theory of vision (in *Optics*)
- Rays emitted *by the eye* traveled along a straight line and formed a cone reaching the scene
- Work set basis for **perspective**

965-1040 CE: Ibn al-Haytham ("Alhazen")
- *Book of Optics* (1028-1038 CE): light rays bounce off objects in all directions and become visible when they reach the eye perpendicularly
Light travels along straight lines – coming from where?

325 BCE: Euclid, a first mathematical theory of vision (in *Optics*)

- Rays emitted by the eye traveled along a straight line and formed a cone reaching the scene
- Work set basis for perspective

965-1040 CE: Ibn al-Haytham (“Alhazen”)

- *Book of Optics* (1028-1038 CE): light rays bounce off objects in all directions and become visible when they reach the eye perpendicularly

From “Total Internal Reflection” – VAN 2020, https://vision.mit.edu/tir
The general rule determining the ideas of vision that are formed whenever an impression is made on the eye, is that *such objects are always imagined as being present in the field of vision as would have to be there in order to produce the same impression on the nervous mechanism, the eyes being used under ordinary normal conditions.*

*Treatise on Physiological Optics* (1866)

- The mind makes perceptions out of sensations finding representations of the object most likely to explain the sensory input.
- Related to Bayesian methods for computer vision inference.
The human visual system recovers scenes that are the *cause* of images.
The human visual system recovers scenes that are the cause of images

Roger Shephard’s Turning Tables (1990)
The human visual system recovers scenes that are the *cause* of images

Perception as inference and explanation, not as measurement
The human visual system recovers scenes that are the *cause* of images.

*Hoarfrost at Ennery*, Camille Pissarro. Oil on canvas, 1873.
The human visual system recovers scenes that are the *cause* of images

Helmholtz 1871 lecture

“We must look upon artists as persons whose observation of sensuous impressions is particularly vivid and accurate, and whose memory for these images is particularly true ... The study of works of art will throw great light on the question as to which elements and relations of our visual impressions are most predominant in determining our conception of what is seen.” (Helmholtz, 1995) (p.280), quoted in (Hyman, 2010).

These scrape marks are sufficient to convey the appearance of a heavily frosted field.

*Hoarfrost at Ennery*, Camille Pissarro. Oil on canvas, 1873
The human visual system recovers scenes that are the *cause* of images – how?

Origins of Structure in Perception: this will take us a semester to answer!

- Low level vision
- Motion and depth
- Color
- Recognition
- Association

And in parallel:

“We must look upon artists as persons whose observation of sensuous impressions is particularly vivid and accurate, and whose memory for these images is particularly true ... The study of works of art will throw great light on the question as to which elements and relations of our visual impressions are most predominant in determining our conception of what is seen.” (Helmholtz, 1995) (p.280), quoted in (Hyman, 2010).
Perception as an inverse problem

Pizlo 2000

“perception is about inferring the properties of the distal stimulus X given the proximal stimulus Y”

\[ Y = AX \]

where
- \( Y \) is the proximal stimulus (e.g. retinal image)
- \( X \) is the distal stimulus (e.g. 3D object)
- \( A \) is a perspective mapping (linear transformation)

\[ X = A^{-1}Y \]

ill-posed, determining a unique \( A^{-1} \) is difficult

many different things could have created the same retinal image (e.g. of a cube)

Less clear why binocular reconstruction of a 3D scene is ill-posed (e.g. Gibson: observer knows position of one eye relative to the other, reconstruction is well-posed and well-conditioned)

Problem: perceptual and motor noise
- Rays never actually intersect
- One could approximate the intersection, but the approximation is very unstable (small variations in the noise cause large variations in the solution)
Perception as an inverse problem
Pizlo 2000

“perception is about inferring the properties of the distal stimulus $X$ given the proximal stimulus $Y$”

$$Y = AX$$

where
$Y$ is the proximal stimulus (e.g. retinal image)
$X$ is the distal stimulus (e.g. 3D object)
$A$ is a perspective mapping (linear transformation)

many different things could have created the same retinal image (e.g. of a cube)

$$X = A^{-1}Y$$

ill-posed, determining a unique $A^{-1}$ is difficult

to solve: $X = A^{-1}Y$

regularized linear regression:

$$E_1 = \| AX - Y_X \|^2 + \lambda \| P_X \|^2$$

Bayes’ rule:

$$p(X|Y_X) = \frac{p(Y_X|X)p(X)}{p(Y_X)} = -\log p(X|Y_X) = -\log p(Y_X|X) - \log p(X)$$
Predictive Coding

“the brain is in the game of optimizing neuronal dynamics and connectivity to maximize the evidence for its model of the world” (Friston 2018, “Does Predictive Coding Have a Future?”)

- neuronal representations in higher/deeper levels of hierarchies generate predictions about representations in lower levels

- descending predictions are compared with lower-level representations to form a prediction error, which is passed back up the hierarchy to update higher-level representations

- The recurrent exchange of signals between adjacent levels encodes a generative model – generating predictions of sensations that can be compared with actual sensory samples
Bayesian Models of Object Perception

“It seems reasonable to assume that an effective perceptual system should obtain and use knowledge of the constraints that characterize the distal stimuli to 'make up' for the information that has been lost in the transformation from the distal to proximal stimulus.” (Pizlo 2000)

adapted from Kersten & Yuille (2003)
adapted from Kersten & Yuille (2003)
image data, \( I \)

object descriptions, \( S \)

telling the robot what they are seeing. If you could see the world through a robot’s eyes, it would look not like a movie picture decorated with crosshairs but something like this:

<table>
<thead>
<tr>
<th>225</th>
<th>221</th>
<th>216</th>
<th>219</th>
<th>214</th>
<th>219</th>
<th>214</th>
<th>207</th>
<th>218</th>
<th>219</th>
<th>220</th>
<th>207</th>
<th>155</th>
<th>136</th>
<th>135</th>
</tr>
</thead>
<tbody>
<tr>
<td>213</td>
<td>206</td>
<td>213</td>
<td>223</td>
<td>208</td>
<td>217</td>
<td>232</td>
<td>223</td>
<td>221</td>
<td>233</td>
<td>216</td>
<td>195</td>
<td>156</td>
<td>141</td>
<td>130</td>
</tr>
<tr>
<td>206</td>
<td>217</td>
<td>210</td>
<td>216</td>
<td>224</td>
<td>223</td>
<td>228</td>
<td>230</td>
<td>234</td>
<td>216</td>
<td>207</td>
<td>157</td>
<td>136</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>211</td>
<td>213</td>
<td>221</td>
<td>223</td>
<td>220</td>
<td>222</td>
<td>237</td>
<td>216</td>
<td>219</td>
<td>220</td>
<td>176</td>
<td>149</td>
<td>137</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>221</td>
<td>229</td>
<td>218</td>
<td>230</td>
<td>228</td>
<td>214</td>
<td>213</td>
<td>209</td>
<td>198</td>
<td>224</td>
<td>161</td>
<td>140</td>
<td>133</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>219</td>
<td>224</td>
<td>220</td>
<td>219-215</td>
<td>215</td>
<td>206</td>
<td>206</td>
<td>221</td>
<td>219</td>
<td>143</td>
<td>133</td>
<td>131</td>
<td></td>
<td></td>
</tr>
<tr>
<td>221</td>
<td>215</td>
<td>211</td>
<td>214</td>
<td>220</td>
<td>218</td>
<td>212</td>
<td>212</td>
<td>212</td>
<td>204</td>
<td>148</td>
<td>141</td>
<td>131</td>
<td></td>
<td></td>
</tr>
<tr>
<td>214</td>
<td>211</td>
<td>218</td>
<td>214</td>
<td>212</td>
<td>216</td>
<td>212</td>
<td>206</td>
<td>206</td>
<td>141</td>
<td>133</td>
<td>131</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>221</td>
<td>208</td>
<td>223</td>
<td>213</td>
<td>216</td>
<td>226</td>
<td>216</td>
<td>225</td>
<td>220</td>
<td>209</td>
<td>143</td>
<td>141</td>
<td>142</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>211</td>
<td>208</td>
<td>223</td>
<td>213</td>
<td>216</td>
<td>224</td>
<td>223</td>
<td>220</td>
<td>219</td>
<td>220</td>
<td>141</td>
<td>133</td>
<td>132</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>204</td>
<td>219</td>
<td>215</td>
<td>217</td>
<td>224</td>
<td>232</td>
<td>241</td>
<td>240</td>
<td>211</td>
<td>150</td>
<td>139</td>
<td>128</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>204</td>
<td>206</td>
<td>208</td>
<td>205</td>
<td>233</td>
<td>241</td>
<td>241</td>
<td>252</td>
<td>242</td>
<td>192</td>
<td>151</td>
<td>141</td>
<td>133</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>205</td>
<td>201</td>
<td>216</td>
<td>232</td>
<td>246</td>
<td>255</td>
<td>266</td>
<td>251</td>
<td>210</td>
<td>149</td>
<td>141</td>
<td>132</td>
<td>126</td>
<td></td>
</tr>
<tr>
<td>191</td>
<td>194</td>
<td>209</td>
<td>238</td>
<td>245</td>
<td>255</td>
<td>249</td>
<td>235</td>
<td>238</td>
<td>197</td>
<td>146</td>
<td>139</td>
<td>130</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>189</td>
<td>199</td>
<td>200</td>
<td>227</td>
<td>239</td>
<td>237</td>
<td>235</td>
<td>236</td>
<td>247</td>
<td>219</td>
<td>145</td>
<td>142</td>
<td>142</td>
<td>133</td>
<td></td>
</tr>
<tr>
<td>198</td>
<td>196</td>
<td>209</td>
<td>211</td>
<td>210</td>
<td>215</td>
<td>236</td>
<td>240</td>
<td>232</td>
<td>177</td>
<td>142</td>
<td>137</td>
<td>135</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>198</td>
<td>203</td>
<td>205</td>
<td>208</td>
<td>211</td>
<td>224</td>
<td>226</td>
<td>240</td>
<td>210</td>
<td>160</td>
<td>139</td>
<td>132</td>
<td>129</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>216</td>
<td>209</td>
<td>214</td>
<td>220</td>
<td>210</td>
<td>231</td>
<td>245</td>
<td>219</td>
<td>169</td>
<td>143</td>
<td>148</td>
<td>129</td>
<td>128</td>
<td>136</td>
<td></td>
</tr>
<tr>
<td>211</td>
<td>210</td>
<td>217</td>
<td>218</td>
<td>214</td>
<td>224</td>
<td>221</td>
<td>212</td>
<td>162</td>
<td>140</td>
<td>139</td>
<td>129</td>
<td>133</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>215</td>
<td>210</td>
<td>216</td>
<td>216</td>
<td>209</td>
<td>202</td>
<td>240</td>
<td>200</td>
<td>156</td>
<td>139</td>
<td>131</td>
<td>129</td>
<td>139</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td>195</td>
<td>220</td>
<td>211</td>
<td>208</td>
<td>205</td>
<td>209</td>
<td>209</td>
<td>240</td>
<td>217</td>
<td>154</td>
<td>141</td>
<td>127</td>
<td>130</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>229</td>
<td>224</td>
<td>212</td>
<td>214</td>
<td>220</td>
<td>229</td>
<td>234</td>
<td>208</td>
<td>151</td>
<td>145</td>
<td>128</td>
<td>128</td>
<td>142</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>252</td>
<td>224</td>
<td>222</td>
<td>224</td>
<td>223</td>
<td>244</td>
<td>228</td>
<td>213</td>
<td>143</td>
<td>141</td>
<td>135</td>
<td>128</td>
<td>131</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>255</td>
<td>235</td>
<td>230</td>
<td>249</td>
<td>253</td>
<td>240</td>
<td>228</td>
<td>193</td>
<td>147</td>
<td>139</td>
<td>132</td>
<td>128</td>
<td>136</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>245</td>
<td>238</td>
<td>245</td>
<td>246</td>
<td>235</td>
<td>235</td>
<td>190</td>
<td>139</td>
<td>136</td>
<td>134</td>
<td>135</td>
<td>126</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>238</td>
<td>233</td>
<td>232</td>
<td>235</td>
<td>255</td>
<td>246</td>
<td>168</td>
<td>156</td>
<td>144</td>
<td>129</td>
<td>127</td>
<td>136</td>
<td>134</td>
<td></td>
</tr>
</tbody>
</table>

adapted from Kersten & Yuille (2003)
prior probability

adapted from Kersten & Yuille (2003)
unlikely                             likely

prior probability

adapted from Kersten & Yuille (2003)

(Mental) image generation
eliminate everything in world model that could not generate the image (likelihood) prior probability then organizes that narrower possibility space.
(Mental) image generation

object descriptions, S

likely

unlikely

prior probability

2D projection

Perceived 3D shape

adapted from Kersten & Yuille (2003)
(Mental) image generation

object descriptions, S

likely
unlikely

prior probability

2D projection

Perceived 3D shape

adapted from Kersten & Yuille (2003)
Visual system embodies principles of ecological optics

- Lighting from above, shadows, reflection, ...

“evolutionary internalized regularities”
Visual system embodies principles of ecological optics

- Interactions of shape and illumination

**Figure 1.** The experimental stimulus. **a.** The chromatic Mach card. The bounding contours of the surfaces, leading to the different shape percepts, which in turn determine one’s percept of the reflectance patterns on the front faces of the two surfaces.

**Figure 3.** Shows the results from the experiment. The 5.8% average increase in luminance needed to make the target patch
Learned (?) Models in Art

Koenderink (2014)

Possible renderings of a common shape are infinite but no problem!
The Beholder’s Share

1909-2001: Ernst Gombrich, Austrian art historian

“draws [the beholder] into the magic circle of creation and allows him to experience something of the thrill of ‘making’ which had once been the privilege of the artist” (Gombrich, 1961) (p. 202).
The Beholder’s Share

Perceptual Metamerization:

• Stimuli that differ physically but look the same (from a particular distance and point of fixation)
• Creation requires a model of perceptual processing
• For example, spatial pooling, both in successive ventral stream areas, and with eccentricity, induces an irretrievable loss of information:
The Beholder’s Share

1909-2001: Ernst Gombrich, Austrian art historian

“draws [the beholder] into the magic circle of creation and allows him to experience something of the thrill of ‘making’ which had once been the privilege of the artist” (Gombrich, 1961) (p. 202).

Koenderink 2012:

QUESTION: Is perception like a bundle of possible visual worlds (then the observer apparently abstains from a final – precognitive – decision) or is it like a single visual world (the observer really “sticks the neck out”)?

In other words, do observers resolve ambiguities when not required for an action or decision?
Minor White \textit{Windowsill Daydreaming}
July 1958
**QUESTION:** Is perception like a bundle of possible visual worlds (then the observer apparently abstains from a final – precognitive – decision) or is it like a single visual world (the observer really “sticks the neck out”)?

---

I think that in a great many cases perceptions are more of the many visual worlds variety than like the single guess. You don’t notice this in the laboratory because most psychophysical methods *force a unique response*. They simulate the decisions–in–action of daily life. That you don’t notice the essential ambiguity of perception in real life is most likely due to the fact that you don’t need to take decisions on issues on which no actions will be taken anyway. That the many visual worlds option is indeed likely is suggested by the fact that a change of psychophysical method or task often leads to distinctly different results. This is not to say that observers actually entertain many visual worlds interpretations explicitly, but merely that they don’t necessarily resolve ambiguities when this is not specifically required for some action or decision.

*Multiple Worlds* (2012)

---

Consciousness: facing the consequences of decisions-in-action? (Schrodinger)
I think that in a great many cases perceptions are more of the many visual worlds variety than like the single guess. You don’t notice this in the laboratory because most psychophysical methods force a unique response. They simulate the decisions—in-action of daily life. That you don’t notice the essential ambiguity of perception in real life is most likely due to the fact that you don’t need to take decisions on issues on which no actions will be taken anyway. That the many visual worlds option is indeed likely is suggested by the fact that a change of psychophysical method or task often leads to distinctly different results. This is not to say that observers actually entertain many visual worlds interpretations explicitly, but merely that they don’t necessarily resolve ambiguities when this is not specifically required for some action or decision.

Multiple Worlds (2012)

Consciousness: facing the consequences of decisions-in-action? (Schroedinger)

Do decisions have consequences when looking at art?

Non-representational art?

Kandinsky (1912)
I think that in a great many cases perceptions are more of the many visual worlds variety than like the single guess. You don’t notice this in the laboratory because most psychophysical methods force a unique response. They simulate the decisions–in–action of daily life. That you don’t notice the essential ambiguity of perception in real life is most likely due to the fact that you don’t need to take decisions on issues on which no actions will be taken anyway. That the many visual worlds option is indeed likely is suggested by the fact that a change of psychophysical method or task often leads to distinctly different results. This is not to say that observers actually entertain many visual worlds interpretations explicitly, but merely that they don’t necessarily resolve ambiguities when this is not specifically required for some action or decision.

Multiple Worlds (2012)

Consciousness: facing the consequences of decisions-in-action? (Schrodinger)

What is the nature of your visual awareness? A single visual world?

Joseph Mallord William Turner (Interior at Petworth c. 1837)