

# ¿Es el mundo como lo vemos?

## Definiendo la pregunta

Hace aproximadamente 2400 años se publicó en la antigua Grecia la *República* de Platón, conteniendo su famosa alegoría de la cueva. Esta alegoría describe un pueblo de personas que viven adentro de una cueva oscura, en la que lo único visible son las sombras que se proyectan desde la entrada de la cueva sobre una de sus paredes. Al no ver otra cosa, estas personas piensan que el mundo está compuesto de sombras, y no pueden concebir el mundo como lo percibimos nosotros. El rol de esta alegoría es que nos cuestionemos la posibilidad de que nos encontremos en una situación similar a la de estas personas, y de que nos estemos perdiendo una parte importante del mundo. Así Platón da lugar a la pregunta central de este capítulo: ¿Es el mundo como lo vemos? A pesar de su larga trayectoria, la filosofía aún no ha abandonado a esta pregunta, con otros pintorescos ejemplos como el demonio de Descartes del 1641, o los cerebros en frascos discutidos recientemente por Putnam, y en sus diversas variantes ha sido una de las principales preguntas guía en la historia de la filosofía.

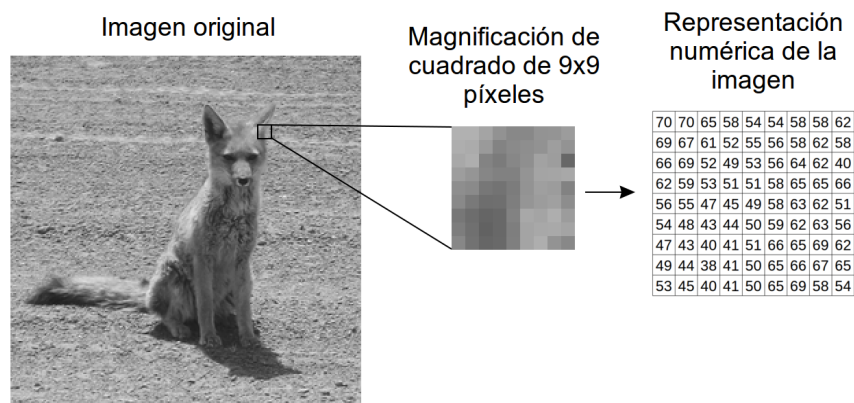
Aunque esta pregunta parece ser simple, en realidad puede ser interpretada y analizada de muchas formas distintas. Por eso, para acercarnos a responderla debemos aclarar a qué nos referimos con las palabras “ver” y “mundo”. Por ejemplo, si tomamos “ver” como todo lo que podemos conocer con nuestros sentidos y a través de la ciencia, nos estaríamos preguntando si hay un “mundo” que nos escapa y que no podemos conocer por ningún método (o si vivimos en algún tipo de cueva de la que no podemos escapar). Este tipo de análisis se encontraría en el campo de la filosofía. En cambio, si tomamos “ver” como lo que podemos percibir con nuestros sentidos, y “mundo” como todo el mundo físico que conocemos, entonces nos encontramos ante una pregunta para la física. Esta pregunta tiene una clara respuesta: el mundo no es como lo vemos. Sabemos que hay partículas subatómicas, materia oscura, campos magnéticos, y muchas otras entidades físicas que escapan a nuestros órganos sensoriales. En contraste con esos posibles enfoques, en este capítulo analizaremos la pregunta desde un punto de vista biológico y neurocientífico. La variante de la pregunta que estudiaremos puede entonces plantearse así: ¿Lo que vemos con nuestros sentidos, es una imagen fiel del mundo a nuestra escala física?

## ¿Qué es ver? Generalidades del sistema visual

Antes de comenzar a analizar qué dice la biología, es necesario reflexionar sobre qué es ver. Nuestro sistema visual tiene una capacidad y una complejidad increíbles que se traducen en una excelente habilidad para ver. Pero por esto, ver nos resulta muy natural y simple, dificultándonos apreciar la complejidad de la tarea. Para ganar intuición sobre lo complicado que es ver, utilizaremos un sistema simple: un robot con visión.

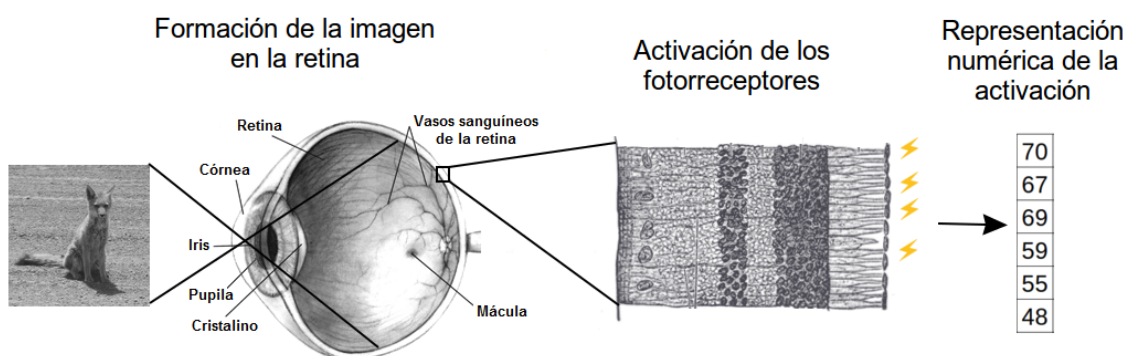
Supongamos que tenemos un pequeño robot y queremos hacer que “vea”. Lo primero que debemos hacer es darle un detector que reciba la luz, o el equivalente a un ojo. Este puede ser simplemente una cámara digital de video. Esta cámara recibe la luz del ambiente y la transforma en una matriz de números indicando la intensidad de la luz en cada región de la escena, lo que constituye una imagen digital (en esta matriz cada número representa un píxel, y su valor indica la intensidad del mismo, como se muestra en la Figura 1). Pero solamente tomar imágenes no alcanza, ver implica utilizar las mismas para extraer información relevante del mundo (ej. qué objeto tengo adelante, a qué distancia está, que tamaño tiene, etc). Entonces debemos agregarle al robot una computadora que reciba esas matrices de números y un programa que las procese para extraer esta información. No es difícil darse cuenta la complejidad que implica generar un programa que le permita al robot convertir esas matrices de números en evaluaciones como “tengo una mesa de mármol de mediana altura a 2 metros en frente mío”. Por ejemplo, notemos que matriz numérica (es decir, la imagen cruda) puede

cambiar radicalmente si modificamos la fuente de iluminación, movemos un poco al robot, cambiamos el fondo de la imagen, o rotamos la mesa, y el programa se enfrenta a la dificultad de dar el mismo resultado ante todos esos cambios “triviales”. También puede verse esta dificultad en la ausencia de autos que se manejen solos y de robots con sistema visual, que denotan que la ciencia y la ingeniería aún no han podido resolver estos problemas.



*Figura 1: Una imagen digital puede tomarse como una matriz de números. Un programa de computadora que analice una imagen debe extraer la información de esta matriz de números.*

Lo interesante del ejemplo del robot es que nuestro sistema visual no es fundamentalmente diferente (ver Figura 2). La luz entra a nuestros ojos y forma una imagen sobre la retina (que se encuentra en la parte de atrás del ojo). La retina está tapizada de células especializadas sensibles a la luz, llamadas fotorreceptores. Los fotorreceptores están activos cuando no reciben luz, y se inactivan al ser iluminados. De esa forma, un fotorreceptor con mayor activación indica una región oscura de la imagen, y uno con mayor activación indica una región más clara, representando así la imagen mediante actividad neuronal (este mecanismo de que mayor actividad indica menos luz es algo anti-intuitiva pero es como el sistema funciona). Notemos que esto es similar a lo que describimos anteriormente para la cámara y los píxeles, y que podemos pensar en cada fotorreceptor como un píxel individual, y en su activación como el valor numérico en la matriz que es la imagen.



*Figura 2: Puede pensarse que el ojo hace una representación numérica de la imagen, si tomamos la activación de cada fotorreceptor como un número sobre el que tiene que operar el cerebro.*

Luego, al igual que para el robot, la imagen debe ser procesada para extraer la información relevante. En nuestro sistema visual, después de un procesamiento inicial en la retina, el nervio óptico transmite la información visual al cerebro, donde tenemos decenas de áreas corticales dedicadas a extraer la información sobre el mundo. Estas áreas se encuentran densamente interconectadas, y su funcionamiento aún no se entiende, pero ellas llevan a cabo el procesamiento que nos permite reconocer una cara, estimar una distancia para tirar un proyectil, o elegir la

manzana que más nos gusta de un cajón. Este procesamiento ocurre de forma subconsciente y no tenemos acceso al mismo, pero es importante tener en cuenta que “ver” constituye un complejo procesamiento e interpretación de la imagen por parte del cerebro.

Ya establecida la complejidad que hay detrás de nuestra capacidad de ver, estamos en condiciones de comenzar a responder la pregunta, es decir, si el procesamiento que nuestro cerebro hace de la imagen resulta en una descripción fiel del mundo que nos rodea. Para ello analizaremos la pregunta desde tres perspectivas: 1) la perspectiva evolutiva, 2) la perspectiva del costo de procesamiento de la información y 3) la perspectiva de la ambigüedad en los datos.

### **Perspectiva evolutiva: El sistema visual no evolucionó para ver el mundo como es**

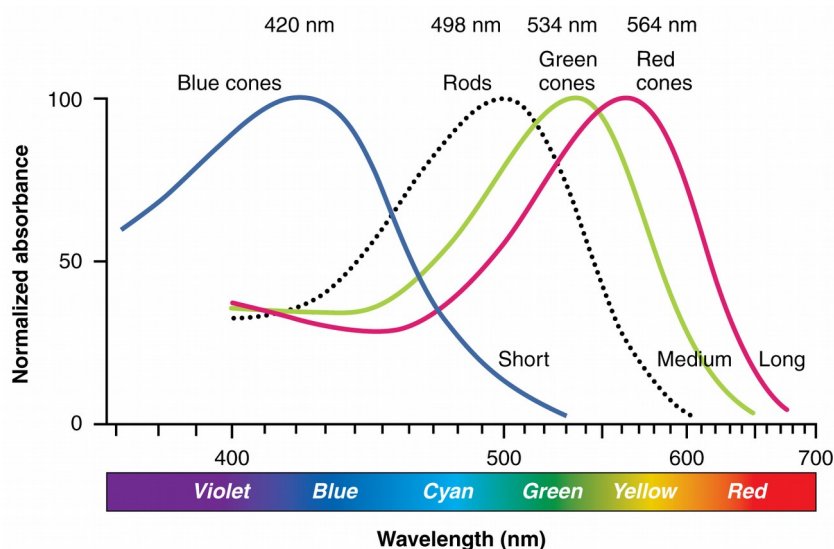
Al igual que el resto de los sistemas y procesos biológicos que componen nuestros organismos, el sistema visual evolucionó por su contribución a favorecer nuestro éxito reproductivo. Es decir, el mismo no evolucionó necesariamente para representar de forma fiel al ambiente, sino porque permite guiar acciones útiles para la supervivencia y la reproducción. A continuación, discutiremos dos ejemplos de cómo la evolución moldea nuestra forma de ver al mundo.

Un ejemplo interesante es el de nuestra visión a color. La luz es una onda electromagnética, y una onda está caracterizada en parte por su longitud de onda. La visión a color consiste en la capacidad de distinguir entre luces con diferentes longitudes de onda. ¿Cómo se logra esto? Para comenzar, un determinado fotorreceptor tiene *preferencia* por determinadas longitudes de onda, que pueden modificar su activación con mayor facilidad (Figura 3). Pero a pesar de esta preferencia, un único fotorreceptor no permite distinguir entre diferentes longitudes de onda porque su activación depende también de la intensidad de la luz. Esto hace que puede activarse lo mismo por una luz de longitud de onda no preferencial pero intensa, o por una luz con su longitud de onda preferida pero tenue, y así no permite distinguir entre ellas. En cambio, dos fotorreceptores con diferentes preferencias de longitud de onda sí permiten discriminar entre las mismas, ya que a partir de sus dos activaciones se puede distinguir la contribución de la intensidad de la luz y de su longitud de onda. Agregando más tipos de fotorreceptores pueden discriminarse mejor las longitudes de onda, mejorando la visión a color, y también pueden abarcarse más longitudes de onda, ampliando el espectro de los colores visibles.

Nuestra visión a color actual tuvo un origen curioso que ilustra algunos puntos interesantes. Durante su evolución, el subfilo de los vertebrados llegó a acumular 4 tipos de fotorreceptores de color distintos, permitiendo una determinada capacidad de visión a color. Pero en algún momento, la clase de los mamíferos perdió 2 de estos fotorreceptores, quedando con una visión a color reducida (la visión a color con 2 fotorreceptores se llama dicromatismo). Sin embargo, nosotros los primates luego adquirimos un nuevo fotorreceptor más, quedando con un total de 3 (y visión a color tricromática). Este fotorreceptor cae sobre el medio de nuestro espectro visible, y permite discriminar más fácilmente el verde de otros colores. Pero, ¿porqué adquirimos un nuevo fotorreceptor? La principal hipótesis dice que nuestra visión tricromática nos es particularmente favorable porque permite discriminar mejor la vegetación, ayudándonos por ejemplo a percibir las frutas que eran parte de nuestra dieta sobre el fondo verde de los árboles.

Esta historia evolutiva de la visión a color muestra cómo nuestros sistemas sensoriales están guiados por las presiones evolutivas, y no necesariamente llevan a percibir el mundo cada vez mejor. Aparentemente, las presiones evolutivas derivadas de nuestro ambiente y dieta resultaron en que adquiramos el tricromatismo, mejorando nuestra percepción a color con respecto a la mayoría de los mamíferos. Sin embargo, en algún momento nuestra percepción a color también se vió reducida en un proceso evolutivo por el cual perdimos dos fotorreceptores. Este evento en particular muestra cómo, en su evolución, nuestros sistemas sensoriales no necesariamente tienden a

representar el mundo cada vez mejor. De esa forma podemos cuestionarnos, ¿cuántos elementos del mundo han dejado de lado nuestros sistemas perceptuales durante la evolución?



*Figura 3: En la gráfica se muestran en las líneas sólidas las preferencias de longitud de onda de nuestros tres fotorreceptores de color. Cada uno se activa con diferentes longitudes de onda. Esto permite desambiguar qué longitud de onda llega a la retina. Otras especies pueden tener más o menos fotorreceptores, permitiéndoles ver más o menos colores respectivamente.*

Otro ejemplo interesante de la relación entre evolución y percepción es el estudio del sistema visual de la rana presentado en “*What the Frog’s Eye Tells the Frog’s Brain*” de Lettvin et.al, publicado en 1959. En este artículo los autores registran la actividad de las fibras en el nervio óptico (que va del ojo al cerebro) de la rana, a la vez que le muestran a la misma diversos estímulos visuales. Así buscan comprender qué información visual llega al cerebro de la rana. Pero en este artículo los autores también prestan atención al comportamiento natural del organismo. La visión es una herramienta importante para la rana, la usa para cazar y para escapar de depredadores. Pero la rana parece no percibir los elementos estáticos del mundo que la rodea: la comida no llama su atención si no se mueve. Y su comportamiento predador se ve guiado principalmente por el movimiento y el tamaño de los objetos visuales: saltará a capturar cualquier objeto pequeño que se mueva como un insecto. De esa forma, su comportamiento visual parece limitado.

Los hallazgos de los autores sobre qué información transmite el ojo de la rana a su cerebro se ajusta muy bien al comportamiento que despliega. Describen 4 tipos de fibras en el nervio óptico, cada una informando sobre ciertos patrones visuales específicos, por lo que el cerebro recibe información ya preprocesada, donde se seleccionó una parte de la información y se desechó otra. Por ejemplo, uno de los tipos de fibras descritos por los autores se activa cuando en la zona del campo visual que representa (lo que se llama el campo receptivo de la fibra) aparece un patrón de luz pequeño que se detiene y se mueve intermitentemente. Esto puede explicar porqué el comportamiento visual de la rana es tan limitado: su cerebro sólo recibe algunos patrones específicos del entorno y no otros (ej. los elementos estáticos de la escena). Sin embargo, aunque a primera vista puede parecernos limitada, la información que se transmite es suficiente para generar los comportamientos que le permiten vivir a la rana. Con respecto a la fibra recién descrita, los autores se preguntan: “¿Puede uno describir un mejor sistema para detectar un insecto accesible?”. Entonces, aunque el mundo visual de la rana es limitado (para ella el mundo visual se compone de algunos patrones simples como puntos moviéndose), es suficiente para guiar los comportamientos esenciales de la rana, y se

ajustan bien a sus necesidades ecológicas (es fácil engañar a una rana en el laboratorio con algo que parezca un insecto, pero ¿cuántos objetos que no son insectos tienen ese tamaño y se mueven como tales en su ambiente natural?). No obstante, es fácil notar una similitud entre el caso de la rana y la alegoría de la cueva, alimentando nuestras dudas sobre nuestra propia percepción.

Pero, siendo que nuestro sistema visual es de una complejidad mucho mayor que el de la rana, podemos preguntarnos ¿en qué medida sufrimos de limitantes tan grandes como el de la rana? Lo descrito para la rana aplica en parte a nuestro sistema visual: el estímulo visual se procesa ya incluso en la retina, y en cada área visual de la corteza se extraen patrones específicos de estímulo que son los que pasan hacia las áreas superiores. Así, se seleccionan algunos patrones y parte de la información pareciera perderse. Sin embargo, hay razones para pensar que nuestro sistema visual no se vea tan afectado como el de la rana. Por ejemplo, hay motivos para pensar que los patrones que se “detectan” en las etapas tempranas de nuestro sistema visual (estructuras tipo “puntos” en la retina y el tálamo, estructuras orientadas tipo “rayas” en la corteza visual primaria) se encuentran entre los patrones que mejor describen nuestro entorno visual. Esto puede apreciarse, por ejemplo, con técnicas de análisis de patrones y aprendizaje automático. En este tipo de análisis se entrenan modelos computacionales para aprender las mejores estrategias de procesamiento para llevar a cabo diferentes tareas visuales como reconocer objetos o navegar el mundo. Los modelos generados luego del aprendizaje muchas veces desarrollan estrategias similares a las de nuestro sistema visual (y el de muchos otros mamíferos complejos), indicando que quizás estemos capturando una gran parte de la información disponible en las imágenes. Esta observación también se condice con un argumento conceptual, respaldado por ejemplo por el filósofo Daniel Dennett: los mamíferos superiores tenemos un comportamiento sumamente complejo y adaptable, que requiere una representación fiel del mundo para poder resolver los problemas a los que frecuentemente nos encontramos. Así podría ser esperable que la evolución nos haya llevado a ver el mundo aproximadamente como es.

En resumen, nuestro sistema visual fue moldeado por la evolución para percibir aquellos aspectos que son importantes para nuestra supervivencia. Sin embargo, nuestro comportamiento complejo y adaptable pareciera requerir la capacidad de percibir fielmente nuestro entorno, y algunos estudios computacionales apoyan este razonamiento. Aunque hay aspectos del mundo que no percibimos, la perspectiva evolutiva deja abierta la posibilidad a que percibamos el mundo de forma precisa.

### **Perspectiva del costo de representación: Representar al mundo fielmente es caro**

Evidentemente, nuestro cerebro es finito y cuenta con recursos limitados. Capturar y procesar los estímulos visuales es un proceso costoso, y tener recursos finitos es una limitante a la hora de percibir fielmente al mundo. No obstante, la evolución también lleva frecuentemente a el uso eficiente de los recursos y, en el caso de nuestro sistema visual, podría ocurrir que los mismos se administren de forma óptima para percibir correctamente lo que nos es relevante. Un aspecto importante e ilustrativo del uso de recursos por nuestro sistema visual es la separación entre visión central y periférica, que describimos a continuación.

Es evidente para nosotros que la visión juega un rol fundamental en nuestra cognición, lo que puede expresarse diciendo que somos un “animal visual”. En línea con esta impresión, el cerebro parece dedicarle una cantidad importante de recursos al procesamiento visual. Pero estos recursos no se distribuyen uniformemente para procesar todo el campo visual, sino que por el contrario su distribución marca una distinción muy fuerte entre dos componentes de la visión: la visión central y la visión periférica. Estos dos componentes son evidentes para nosotros mediante la introspección, ya que mientras percibimos de forma muy nítida la parte de la escena visual sobre la que posamos la mirada, lo que cae en nuestra visión periférica es mucho menos claro. Un claro ejemplo de esto es que nos resulta imposible hacer algunas tareas (ej. leer) con la visión periférica.

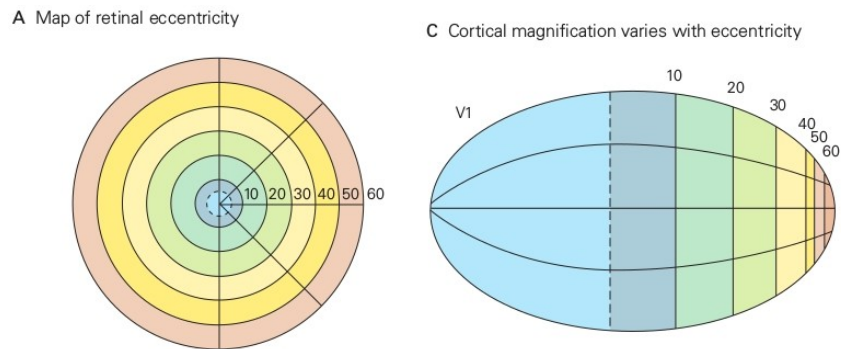


Figura 4: Distribución de recursos corticales a lo largo del campo visual

Esta división entre visión central y periférica nos es tan natural que no la solemos cuestionar, pero ¿porqué vemos con poca nitidez en la periferia? ¿sería posible tener la alta resolución de la visión central en todo el campo visual? La clave para entender esto radica en la distribución de recursos de nuestro sistema visual. Aunque nuestra visión central ocupa una pequeña fracción del campo visual (podemos definirla aproximadamente como el área que abarca nuestro puño cerrado con el brazo extendido, ver Figura 4), la misma utiliza aproximadamente la mitad de los recursos de nuestro sistema visual. Así, aunque la visión periférica ocupa la gran mayoría del campo visual, sólo utiliza la otra mitad de los recursos corticales. Si extrapolamos la relación entre campo visual y área cortical de la visión central, procesar todo nuestro campo visual con la nitidez de la visión central requeriría tener un cerebro mucho más grande que el que tenemos, lo que pareciera ser casi una imposibilidad biológica.

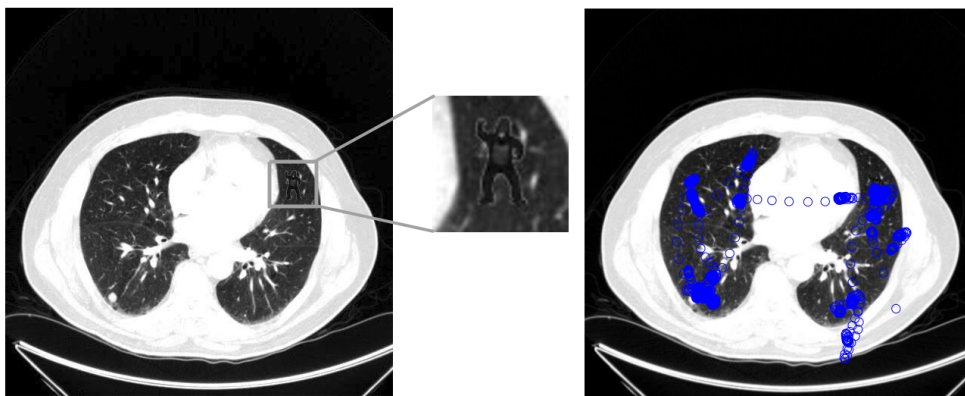
Sin embargo, a pesar de que vemos claramente solo una pequeña parte del campo visual, utilizamos una importante herramienta para aprovechar los recursos disponibles: los movimientos oculares. Estamos constantemente moviendo los ojos (y la cabeza) para capturar con alta precisión los elementos importantes del ambiente. Esto es algo que sabemos hacer tan bien que nos resulta natural y simple, e incluso no notamos los constantes movimientos oculares que realizamos. Pero nuestros ojos nunca están quietos, y en general elegimos (de forma subconsciente) los aspectos más relevantes de las escenas para escanear, permitiéndonos obtener altas cantidades de información con recursos limitados. Así logramos generar un modelo detallado del mundo, y tenemos la experiencia de percibir bien el entorno. Pero, sorprendentemente, a pesar de la eficiencia del sistema, muchas veces sobreestimamos nuestra percepción, como ilustran los siguientes ejemplos.

El fenómeno llamado cegera al cambio, o *change blindness* se demuestra experimentalmente mostrando de forma sucesiva y repetida dos imágenes, que son idénticas a excepción de algún elemento concreto. El experimento consiste en que la participante identifique en qué difieren las imágenes. Lo interesante del fenómeno es que se utilizan imágenes con diferencias grandes, que pensaríamos que son fáciles de identificar, pero al llevar a cabo la tarea nos encontramos con que es mucho más difícil de lo que pensaríamos. Este resultado se contrasta con nuestra impresión subjetiva: creemos ver claramente la imagen que nos muestra y percibir de forma correcta sus elementos más importantes. La realidad es que nos es difícil recordar hasta los elementos más conspicuos de la escena.

Otro ejemplo divertido es el de la cegera inatencional, o *inattention blindness*. Este fenómeno consiste en una llamativa incapacidad para percibir aspectos de las imágenes para los que no estamos atentos, a pesar de que sean muy salientes. Un ejemplo icónico que puede encontrarse en internet buscando "*inattention blindness*" consiste en un video en el que se muestra un grupo de personas haciendo pases con una pelota de basketball, y se plantea la tarea de contar cuantos pases



se hacen. Al terminar el video se revela que un curioso animal hizo una aparición poco disimulada, que sin embargo pasa desapercibida para las personas que se concentran en contar los pases. Otro ejemplo, llamativo por el ojo entrenado que de los participantes, es un experimento en el que se les pidió a radiólogos que realizaran un diagnóstico sobre unas placas de tomografía en las que se había incluido una visible imagen de un gorila (ver Figura 5). A pesar de que los radiólogos examinaron la imagen con detenimiento, y de que casi todos posaron los ojos sobre el gorila (según se midió con la técnica de eye-tracking), cuando se les consultó al final del experimento, la mayoría no había notado al simio oculto. Este último ejemplo muestra además que no alcanza con tener algo en la visión central para poder procesarlo correctamente. La asignación de recursos en el sistema visual no se da solamente en los movimientos oculares, sino también a nivel del procesamiento de la información, priorizando un “camino” de procesamiento sobre otro.



*Figura 5: Imagen de ceguera inatencional. A la izquierda se muestra la tomografía con una ampliación indicando al gorila. A la derecha se indican los movimientos oculares de un participante con los círculos en azul.*

Estos ejemplos ilustran cómo procesar el estímulo visual es costoso, y en un momento determinado percibimos con claridad sólo una pequeña parte de la escena. Una lectura posible de esto es que el mundo no es como lo vemos, ya que hay muchos aspectos notorios del mismo de los que no tomamos conciencia. Pero, ¿es esto algo nuevo? Antes de comenzar esta sección, ya cualquiera de nosotros sabía que hay elementos de la escena que no puedo percibir en un determinado momento, por ejemplo los que están atrás mío. Aunque es interesante cómo nuestra percepción “real” o “objetiva” parece ser peor de lo que creemos, este nuevo descubrimiento no parece cerrar la cuestión, ya que desde un comienzo sabemos que nuestro sistema visual tiene limitantes. Lo que aprendimos es que esas limitantes son mayores de lo que parecen, pero no implican que no pueda percibir bien los elementos específicos del mundo sobre los que presto atención, al igual que no tener ojos en la espalda no significa que no puedo percibir lo que está atrás mío, ya que puedo darme vuelta. Y aunque los ejemplos discutidos nos enseñan mucho sobre el sistema visual, es importante notar que en el mundo real los objetos no suelen desaparecer repentinamente como en el ejemplo del *change blindness*, o que en general nuestras expectativas sobre qué encontrar en un lugar son razonables, haciendo que el *inattention blindness* no sea de una magnitud tan grande. Es decir, podemos exponer las limitantes de las estrategias de asignación de recursos del sistema visual en un laboratorio, pero no es claro en qué medida fallan en el mundo real.

En definitiva, no percibimos todo el mundo al mismo tiempo, pero esto ya era claro. Lo que queda preguntarse ahora es: aquella parte de la escena sobre la que prestamos atención y que vemos con detenimiento, ¿es cómo la vemos?

**Perspectiva de la ambigüedad de los datos: ver requiere “interpretar” al mundo**

El último aspecto que consideraremos para ver en qué medida nuestra percepción se parece al mundo exterior es un aspecto fundamental sobre los aspectos computacionales de la visión: los estímulos visuales deben ser *interpretados* para obtener la información del mundo. Esto es un hecho que viene en parte de que nuestra entrada visual tiene 2 dimensiones (el plano de nuestra retina, al igual que las dos dimensiones de una fotografía), mientras que el mundo exterior tiene 3 dimensiones espaciales. Nuestro sistema visual es muy bueno haciendo interpretaciones de los estímulos visuales, y lo hace de forma subconsciente, dificultándonos apreciar las ambigüedades que tienen. Pero consideremos nuevamente la visión del robot y veamos cómo hay “interpretaciones” a varios niveles.

Uno de los pasos importantes en el procesamiento visual es el de agrupar los píxeles en objetos (proceso que llamamos la segmentación de la imagen). Esto involucra utilizar algún criterio para determinar cuándo dos regiones de la imagen pertenecen a la misma superficie, y así agruparlos. Pero esto no es fácil. Podemos pensar algunas reglas simples para lograrlo (ej. agrupar los píxeles que tienen intensidad y color parecidos y son cercanos), pero también podemos de forma rápida identificar ejemplos en los que esto no se cumpla (ej. una superficie texturada con líneas marcadas, como una prenda de vestir). De hecho, este es un problema abierto en la investigación del procesamiento de imágenes con computadoras. Luego de segmentar la imagen debemos estimar la distribución espacial de estas superficies (ej. a qué distancia se encuentran de nosotros, su orientación espacial, su tamaño, entre otras) y agruparlas en objetos. Continuando con el procesamiento terminamos construyendo una escena en 3 dimensiones de los objetos en nuestro entorno. Todos estos pasos involucran etapas de interpretación (por ejemplo, interpretar si dos “píxeles” son de la misma superficie o no). Así, con una serie considerable de interpretaciones pasamos de una imagen en dos dimensiones en nuestras retinas a los aspectos tridimensionales de la escena. Una imagen famosa que ilustra este proceso es la que se muestra en la Figura 6. Aunque es una imagen compuesta de manchas negras, nuestro cerebro logra en un momento generar un agrupamiento de los elementos, una composición de la estructura de las superficies, y agruparlas en objetos y finalmente en una escena. Por su apariencia “desorganizada” este ejemplo es ilustrativo, ya que es difícil captar qué evidencia nuestro cerebro usa para interpretar ese caos de píxeles. Pero más allá de los detalles del procesamiento, el hecho clave que nos importa aquí es que en las mismas puede haber más de una interpretación posible, lo que ilustran los siguientes ejemplos.



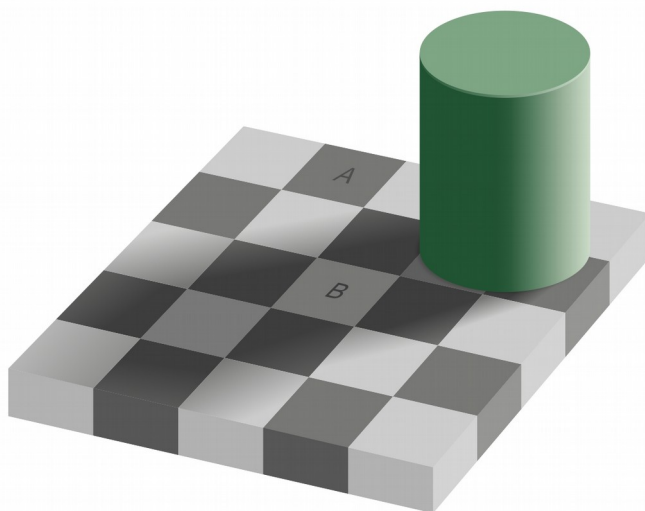
*Figura 6: Intente percibir qué es lo que se muestra en la imagen.*

El cuarto de Ames es una ilusión clásica en la que la percepción que tenemos de una escena está errada. Este es un cuarto con una forma particular: no es rectangular, y tanto el techo como el piso están inclinados, generando una variación en altura a lo largo del cuarto. Para generar la ilusión, una de sus paredes tiene un agujero cuidadosamente ubicado para mirar hacia el interior, y a través del cual percibimos un cuarto cuadrado normal. Además, si ubicamos personas dentro del cuarto en esquinas opuestas, veremos a una de las personas con un tamaño sobrehumano (la que se encuentre



en la parte pequeña del cuarto), y a la otra con un tamaño pequeño (la que se encuentre en la parte extendida del cuarto). En este ejemplo nuestra percepción está errada. Vemos que la percepción que tenemos no es la de la forma real del cuarto. El estímulo visual en este caso (y en todos los otros) es ambiguo. Es compatible tanto con un cuarto cuadrado (como percibimos erróneamente), como con la escena real del cuarto irregular (ya que esta es de hecho la que genera la imagen). De hecho, el estímulo visual es compatible con infinitas escenas posibles. Por ejemplo, podría ser un cuarto gigante con personas gigantes, un cuarto pequeño con personas pequeñas, o podría tener muchas formas distintas con personas de distintos tamaños. La imagen es compatible con todas estas alternativas porque todas podrían generar el mismo patrón de luz en mi retina. Este ejemplo muestra que en la percepción es necesario un paso de interpretar la imagen para adivinar qué escena está detrás de la misma.

Otro ejemplo interesante, que no involucra la distribución espacial de la escena, es la ilusión de Adelson. En la misma se muestra la imagen de una superficie con un patrón como el de un tablero de ajedrez, con cuadrados blancos y negros. Arriba de la misma hay un objeto que proyecta una sombra sobre parte de la superficie, y se marcan un cuadrado negro a la luz y uno blanco a la sombra. Nuestra percepción de la escena es que el cuadrado blanco a la sombra es más claro que el cuadrado negro a la luz, pero en verdad la imagen es generada tal que tengan el mismo tono de gris. Es decir, nuestro sistema visual genera una interpretación de la escena en la que una región de la imagen tiene mayor brillo que otra, pero esta interpretación está equivocada, ya que las dos regiones tienen el mismo brillo.



*Figura 7: Ilusión de Adelson. El cuadrado A tiene el mismo tono de gris que el cuadrado B.*

Estos ejemplos muestran que nuestro sistema visual lleva a cabo interpretaciones de los estímulos visuales para percibir la escena que nos rodea, y que la misma puede estar errada. Pero, ¿cómo llega nuestro cerebro a estas interpretaciones? Aunque esta pregunta aún está abierta, la neurociencia puede darnos algunas respuestas. Primero es importante aclarar que para investigar esto normalmente se estudian estímulos simplificados que apuntan a entender un proceso específico del procesamiento con un tipo específico de información. Algunos ejemplos podrían ser, cómo usamos el color para segmentar las imágenes, o cómo usamos las texturas visuales para estimar distancias. A pesar de la brecha que hay entre este tipo de experimento simplificado y nuestra percepción “natural”, los mismos permiten extraer dos conclusiones relevantes a nuestra pregunta: 1) el sistema visual utiliza un conjunto de “reglas” que permiten construir la interpretación de la imagen y 2) estas reglas no son arbitrarias, sino que están basadas en la estructura del mundo que habitamos.

En el proceso de elegir cual de las escenas posibles es la que genera la imagen el cerebro impone restricciones, o reglas, en la interpretación que ayudan a elegir algunas posibilidades y descartar otras. Estas reglas marcan cómo interpretar determinados elementos de la imagen y se aplican a varios niveles y de forma paralela. Por ejemplo, un fenómeno que se observa con estímulos simples es que nuestra visión agrupa los elementos de la imagen que se encuentran alineados en la dirección de su orientación (que son colineales, ver Figura 7). Por el contrario, elementos que se encuentran alineados en la dirección perpendicular a su orientación, no tienden a ser agrupados. Entonces, utilizando esta regla para interpretar imágenes naturales, nuestro sistema visual puede elegir la interpretación de la escena que agrupe los elementos que son colineales y descartar otras interpretaciones. Otra regla que usa el sistema visual es que tiende a asumir que la fuente de iluminación viene de arriba. Una ambigüedad común es que la misma imagen puede ser generada por una superficie cóncava iluminada desde arriba y por una superficie convexa e iluminada desde abajo. Ante este tipo de ambigüedad por ejemplo nuestro sistema visual tenderá a elegir la escena en que la iluminación viene desde arriba. Entonces, aplicando una gran cantidad de reglas el sistema visual puede ir eligiendo una interpretación sobre otra a diferentes niveles, hasta llegar a una interpretación global de la imagen que recibimos, que será nuestra percepción. Pero, ¿de dónde salen estas reglas? ¿cómo sabemos si las mismas son buenas para percibir el mundo?



*Figura 8: Ejemplo de agrupamiento por orientación en la imagen. Agrupamos de forma rápida e inevitable los elementos colineales a la izquierda. En la derecha Hay elementos con casi la misma organización pero ortogonales que no logramos agrupar. Esto muestra una regla subconsciente que utilizamos para interpretar la imagen.*

La investigación en neurociencia teórica apunta a que estas reglas que usamos para interpretar las imágenes están basadas en las regularidades del mundo que nos rodea. Nuestro mundo tiene una marcada estructura que da lugar a una gran cantidad de regularidades, es decir, patrones que se repiten en las imágenes y que tienen relaciones predecibles con el mundo. Por ejemplo, nuestro ambiente visual está compuesto en su mayoría por objetos sólidos. Debido a las leyes de la física, los objetos generan contornos continuos en las imágenes, que consisten en elementos orientados (bordes) que están alineados en la dirección de su orientación. De esa forma podría explicarse la regla de agrupar los elementos colineales como una consecuencia razonable de la estructura del mundo: los objetos en el mundo tienden a generar elementos colineales, y sería correcto agruparlos para formar una estructura más global. Asimismo, es fácil notar que en su amplia mayoría las fuentes de iluminación en nuestro entorno natural vienen de arriba hacia abajo, por lo que tiene sentido usar una regla que elija las interpretaciones de la escena con esta característica. De esta forma, las reglas que usa nuestro sistema visual estarían asociadas a la estructura del mundo que nos rodea, y ayudarían a elegir las interpretaciones de las imágenes que más se ajusten al mismo.

Pero, si estas reglas se derivan de la estructura del mundo para representarlo correctamente, ¿porqué percibimos de forma errónea el cuarto de Ames o la ilusión de Adelson? La respuesta es que son

estímulos atípicos que violan la estructura normal del mundo. La imagen del cuarto de Ames es generada por una perspectiva determinada sobre un cuarto con una forma irregular con forma muy específica, pero esta es una disposición que difícilmente ocurra en el mundo real. Es mucho más probable que una imagen de ese tipo sea generada por un cuarto cuadrado. De la misma forma, en la ilusión de Ames la imagen es una foto en 2 dimensiones que emula una escena tridimensional con una sombra que cae sobre la superficie. Aunque en la imagen artificial los dos cuadrados tienen el mismo tono de gris, en una escena real con un objeto y una sombra, los cuadrados serían de colores distintos (uno negro y otro blanco), y nuestra interpretación sería correcta (ya que queremos saber el color del cuadrado en sí). Entonces, estas ilusiones que muestran cómo nuestro cerebro construye interpretaciones, y cómo estas pueden estar equivocadas, en verdad se dan por el uso de estímulos rebuscados, que difícilmente encontremos en el mundo real.

En conclusión, las imágenes son fundamentalmente ambiguas, y una imagen determinada es compatible con infinitas escenas posibles. Así percibir requiere obligatoriamente de un proceso de interpretación de la imagen, en la que elegimos una interpretación posible por sobre otras. Pero este proceso no es arbitrario, sino que responde a reglas que se ajustan bien a la estructura del mundo, logrando que en general nuestras interpretaciones sean buenas.

## **Conclusión**

En suma, vimos tres motivos neurobiológicos por los que el mundo podría no ser cómo lo vemos: 1) nuestro sistema visual no evolucionó porque permita ver al mundo como es; sino porque permite guiar acciones útiles; 2) representar al mundo como es resulta muy caro; y 3) los estímulos que recibimos son ambiguos, y requieren una cuota importante de interpretación de nuestra parte. En los tres casos vimos ejemplos que pueden sugerir que el mundo no es como lo vemos. Pero también discutimos razones por las que estas limitaciones no son determinantes: 1) tenemos un comportamiento muy complejo que puede requerir una percepción fiel del mundo (y la investigación computacional parece decir que nuestra percepción es buena); 2) asignamos muy bien nuestros recursos de procesamiento, y aunque parte importante del mundo nos escapa (lo que es inevitable), nos armamos una imagen decente del mismo; y 3) nuestras interpretaciones utilizan reglas que están basadas en la estructura del mundo que nos rodea, proporcionándole una base sólida a las mismas. En definitiva, la respuesta a la pregunta de si el mundo es cómo lo vemos dependerá de la definición precisa de la pregunta, y los estándares que tengamos sobre qué constituye “ver al mundo como es”. Y si genera desilusión no contar con una respuesta concreta al final del capítulo, quizás ayude recordar que a pesar de su aspecto simple, esta pregunta ha dado (y sigue dando) qué discutir durante una parte importante de la historia de las ideas.

Para finalizar, vale la pena destacar que aunque al comienzo del capítulo marcamos una distinción entre la pregunta filosófica y la neurocientífica o cognitiva, la filosofía es una parte integral de las ciencias cognitivas y de la neurociencia. Muchas de las preguntas que estudian estas últimas se basan en conceptos que, al escarbarlos con profundidad, devienen en análisis y discusiones que actualmente pertenecen a la filosofía. Incluso, aunque no lo marcamos explícitamente, muchos puntos que dimos por sentado en este capítulo se alinean con una u otra perspectiva filosófica, y pueden ser duramente criticados desde otras perspectivas filosóficas (por ejemplo, es discutible que los estímulos visuales que recibimos sean realmente ambiguos). Quizás una buena forma de cerrar el capítulo es revisitando a Daniel Dennett (importante filósofo de la ciencia), quien expresa de forma elocuente esta idea: “No hay tal cosa como ciencia libre de filosofía; solamente hay ciencia cuyo bagaje filosófico es llevado a bordo sin examinación”.