

Capítulo 7

¿ES EL MUNDO COMO LO VEMOS?

Daniel Herrera

Laboratorio de Neurociencias, Facultad de Ciencias,
Universidad de la República, Uruguay

HACE APROXIMADAMENTE 2.400 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 dentro 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. Esta alegoría hace que nos cuestionemos la posibilidad de encontrarnos en una situación similar a la de estas personas y de que nuestra percepción nos presente un mundo distorsionado. 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 que, con sus variantes y con otros pintorescos ejemplos como el demonio de Descartes de 1641, 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,

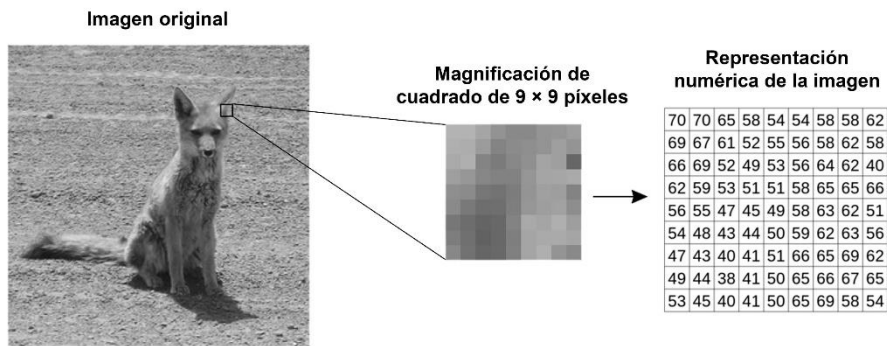
campos magnéticos y muchas otras entidades físicas que escapan a nuestros órganos sensoriales. En contraste con esos posibles enfoques, en este capítulo analizaremos la pregunta desde un punto de vista biológico y neurocientífico. La variante de la pregunta que estudiaremos puede entonces plantearse 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 la facilidad con la que vemos nos dificulta 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 7.1). Pero con tomar imágenes no alcanza. Ver implica utilizar las mismas para extraer información relevante del mundo (*e.g.*, qué objeto tengo adelante; a qué distancia está; que tamaño tiene, etc.). Para ello debemos agregarle al robot una computadora con un programa que procese las matrices de números 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 dos metros enfrente de mí”. Por ejemplo, notemos que la matriz numérica (*i.e.*, 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. El programa se enfrenta a la dificultad de dar el mismo resultado ante todos esos cambios “triviales”. Pueden verse estas dificultades al desarrollar autos que se manejen solos (que suelen usar sensores complementarios y funcionar solo en ambientes controlados o previamente mapeados) y robots con sistema visual que puedan ir más allá de las tareas más básicas. La ciencia y la ingeniería aún no han podido resolver estos problemas.

Figura 7.1



Una imagen digital puede pensarse como una matriz de números. A la izquierda vemos una imagen en blanco y negro y en el centro una ampliación de un área de la misma. A la derecha vemos la misma área de la imagen pero con números entre 0 y 100 que indican el valor de intensidad de cada píxel. Un programa de computadora que “vea” debe tomar esa matriz de números y extraer de ella la información de la escena (e.g., que la misma contiene un zorro).

Lo interesante del ejemplo del robot es que nuestro sistema visual no es fundamentalmente diferente. 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 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 menor activación indica una región más clara, representando así la imagen mediante actividad neuronal. Este sistema mediante el cual se sensa el grado de iluminación puede entonces pensarse como el descrito anteriormente, siendo cada fotorreceptor como un píxel individual y su grado de activación como el valor numérico en la matriz que es la imagen.

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 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.

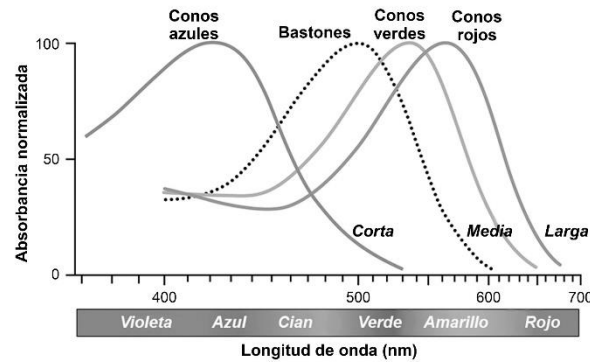
Ya establecida la complejidad que hay detrás de nuestra capacidad de ver, estamos en condiciones de comenzar a responder la pregunta, es decir, si el procesamiento que nuestro cerebro hace de la imagen resulta en una descripción fiel del mundo que nos rodea. Para ello analizaremos la pregunta desde tres perspectivas: 1) la evolutiva; 2) la del costo de procesamiento de la información, y 3) la 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 para contribuir a favorecer nuestro éxito reproductivo. Es decir, en la evolución no fue seleccionado necesariamente el sistema visual que mejor representaba el ambiente, sino aquel que más favorecía la supervivencia y la reproducción. Un ejemplo que ilustra este punto es el de nuestra visión a color, que recorrió un camino interesante para llegar a su estado actual.

La luz es una onda electromagnética y, al igual que otras ondas, está caracterizada en parte por su longitud de onda. La visión a color consiste en la capacidad de distinguir la longitud de onda de la luz que percibimos. ¿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 7.2). 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 pueda tener la misma activación con una luz de longitud de onda no preferencial pero intensa o con una luz de 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.

Figura 7.2



Los seres humanos tenemos tres fotorreceptores distintos para visión a color. Cada uno de estos receptores, llamados conos, tiene su preferencia por diferentes longitudes de onda. En la gráfica se muestra la capacidad de cada cono (nombrados azul, verde y rojo) de absorber las diferentes longitudes de onda (cuyo color se indica en la barra de abajo). Un único cono no permite desambiguar la longitud de onda porque su activación depende del color de la luz y de la intensidad de la misma (e.g., no puede determinar si una mayor activación se debe a un cambio en el color de la luz o a un cambio en su intensidad).

En un momento de la historia, el subfilo de los vertebrados llegó a acumular cuatro tipos de fotorreceptores de color distintos permitiendo una determinada capacidad de visión a color. Pero luego, los mamíferos, en un período de su evolución en que eran animales nocturnos, perdieron dos de estos fotorreceptores quedando con dos fotorreceptores (o visión *dicromática*). Sin embargo, posteriormente los primates adquirimos un nuevo fotorreceptor quedando con un total de tres (y visión a color *tricromática*). Este fotorreceptor cae sobre el medio de nuestro espectro visible y permite discriminar más fácilmente el verde de los otros colores. La principal hipótesis sobre qué llevó a adquirirlo es que fue particularmente favorable para percibir mejor la vegetación, que era parte importante de nuestro ambiente, ayudándonos a distinguir por ejemplo las frutas, que eran parte de nuestra dieta, sobre el fondo verde de los árboles. Estos dos eventos evolutivos muestran cómo la evolución puede llevar a mejoras en los sistemas sensoriales, aunque estos también pueden empeorar en ciertos aspectos si la evolución produce cambios por la relación costo-beneficio de los mismos (ver más adelante). De esta forma, la evolución ha llevado a que actualmente percibamos menos colores que muchas especies de pájaros que aún mantienen sus cuatro fotorreceptores.

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. El experimento llevado a cabo en este artículo consiste en mostrar estímulos visuales a una rana mientras se registra la actividad de las fibras de su nervio óptico (el que va del ojo

al cerebro), para determinar qué información sobre la escena transmiten estas fibras al cerebro. Lo interesante de este artículo es que los autores discuten los resultados en el contexto del comportamiento natural del espécimen, que utiliza la visión para cazar y para escapar de depredadores. Por ejemplo, el comportamiento visual de la rana tiene la peculiaridad de que la misma parece no percibir los elementos estáticos del mundo que la rodea (*e.g.*, la comida no llama su atención si no se mueve). Además, su comportamiento predador se ve guiado principalmente por el movimiento y el tamaño de los objetos visuales: intentará capturar cualquier objeto pequeño que se mueva como un insecto, aunque para nuestro sistema visual parezca muy distinto. Los autores discuten que este comportamiento parece ajustarse muy bien a sus hallazgos sobre la activación de las fibras.

En este estudio se describen cuatro tipos de fibras en el nervio óptico, cada una informando sobre ciertos patrones visuales específicos en el ambiente. Es decir, el cerebro recibe información ya preprocesada en la que se seleccionó una parte de la información (la correspondiente a los patrones específicos) y se desechó otra. Esto puede explicar por qué el comportamiento visual de la rana es tan limitado: su cerebro solo recibe algunos patrones específicos del entorno y no otros (*e.g.*, los elementos estáticos de la escena). Por ejemplo, uno de los tipos de fibras descritos por los autores se activa cuando en la zona del campo visual que representa aparece un patrón de luz pequeño que se detiene y se mueve intermitentemente. Entonces la rana podrá detectar este tipo de evento visual, pero no otros para los que no tenga fibras.

Pero, aunque la percepción de la rana puede parecernos limitada, lo importante es que le permita generar los comportamientos necesarios para sobrevivir y reproducirse. Por ejemplo, con respecto a la fibra descrita anteriormente 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 se compone de algunos patrones simples como puntos moviéndose), es suficiente para sus necesidades ecológicas. Aunque es fácil engañar a una rana en el laboratorio con algo que parezca un insecto, la pregunta relevante es: ¿cuántos objetos que no son insectos tienen ese tamaño y se mueven como tales en su ambiente natural?

Es posible notar una similitud entre el caso de la rana y la alegoría de la cueva, y aunque nuestro sistema visual es de una complejidad mucho mayor, es natural preguntarse ahora ¿en qué medida sufrimos de limitantes tan grandes como las de la rana? De cierta forma nuestro sistema visual lleva a cabo un

procesamiento similar: el estímulo visual se procesa ya incluso en la retina, y en cada área visual de la corteza cerebral se extraen patrones específicos de estímulo que pasan hacia las áreas superiores. Sin embargo, hay razones para pensar que nuestro sistema visual no se ve tan afectado como el de la rana. Por ejemplo, los patrones que se detectan en las etapas tempranas de nuestro sistema visual (estructuras tipo “puntos” en la retina; estructuras orientadas tipo “rayas” en la corteza visual primaria) son los mismos a los que llegan herramientas matemáticas que buscan el mejor método de representar las imágenes. Esto sugiere 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 con 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 parecería 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 de que percibamos el mundo de forma precisa.

Perspectiva del costo de representación: representar al mundo fielmente es caro

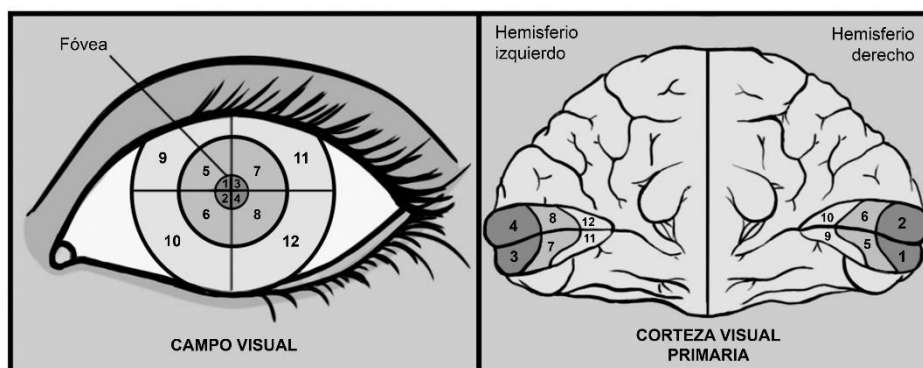
Nuestro cerebro es finito y cuenta con recursos limitados. Capturar y procesar los estímulos visuales es costoso, por lo que nuestra capacidad de percibir fielmente el mundo también es limitada. No obstante, la evolución suele llevar al uso eficiente de los recursos, y ocurre que nuestro sistema visual administra los mismos de forma óptima permitiéndonos 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 que la visión juega un rol fundamental en nuestra cognición, lo que puede expresarse diciendo que somos un “animal visual”. En línea con

esto, el cerebro dedica una cantidad importante de recursos al procesamiento visual. Pero estos recursos no se distribuyen uniformemente para procesar todo el campo visual, sino que su distribución marca una distinción muy fuerte entre dos componentes de la visión: la visión central y la visión periférica. Estos componentes son evidentes para nosotros mediante la introspección: 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 ejemplo evidente de esto es que nos resulta imposible hacer algunas tareas (como leer) con la visión periférica.

Esta división entre visión central y periférica nos es tan natural que no la solemos cuestionar, ¿pero por qué vemos con poca nitidez en la periferia?; ¿sería posible tener la alta resolución de la visión central en todo el campo visual? Una clave para entender esto es que, 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; Figura 7.3), 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, solo 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 parecería ser casi una imposibilidad biológica.

Figura 7.3



Los recursos del sistema visual no se distribuyen de forma homogénea para todo nuestro campo visual. En la imagen se muestra un mapa entre el campo visual (a la izquierda) y la corteza visual primaria (a la derecha). Vemos que casi la mitad de la corteza visual primaria está dedicada a procesar una pequeña fracción del campo visual que se denomina fovea.

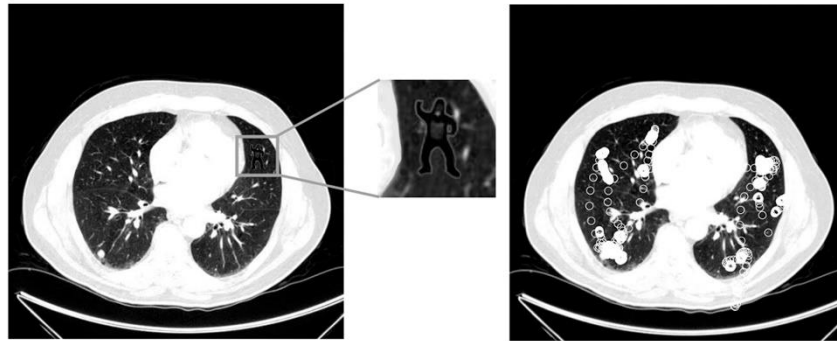
Pero a pesar de que vemos claramente solo una pequeña parte del campo visual, utilizamos una importante herramienta para aprovechar los recursos disponibles: los movimientos oculares. Estamos constantemente moviendo los ojos (y la cabeza) para capturar con alta precisión los elementos importantes del ambiente. Este escaneo de la escena es algo que sabemos hacer tan bien que nos resulta natural y simple, y 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. Con esta asignación eficiente de recursos logramos armar un modelo detallado del entorno. Pero a pesar de la eficiencia del sistema muchas veces sobreestimamos nuestra percepción, como ilustran los siguientes ejemplos.

El fenómeno llamado ceguera 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 (se pueden encontrar varios ejemplos en internet). El experimento consiste en que el participante identifique en qué difieren las imágenes. Lo interesante del fenómeno es que incluso diferencias grandes entre las imágenes son difíciles de identificar. Este resultado se contrasta con nuestra impresión subjetiva: aunque creemos percibir claramente la imagen, la realidad es que nos es difícil recordar hasta sus elementos más conspicuos.

Otro ejemplo es el de la ceguera inatencional (o *inattention blindness*). Este fenómeno consiste en que ciertos elementos de las imágenes a los que no estamos prestando atención pueden pasarnos desapercibidos a pesar de ser muy visibles. Un ejemplo icónico que puede encontrarse en internet consiste en un video en el que se muestra un grupo de personas haciendo pases con una pelota de básquetbol y se plantea la tarea de contar cuántos 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 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 (Figura 7.4). 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 7.4



La ceguera inatencional es un fenómeno en el cual aspectos del campo visual que no esperamos o que no estamos atendiendo pasan desapercibidos. En la figura se muestra una imagen de un experimento clásico en que se insertó un gorila en placas de tomografía, como se muestra en la ampliación. En el experimento se le pidió a radiólogos que evalúen la tomografía sin avisarles de la presencia del gorila. La mayoría de los radiólogos no lo detectó, a pesar de que muchos posaron su mirada sobre el mismo (los círculos blancos a la derecha muestran el recorrido de la mirada de uno de los radiólogos). Tomada de Drew, Vö & Wolfe (2013), *Psychol. Sci.* 9: 1848.

Estos ejemplos ilustran cómo es de costoso procesar el estímulo visual y cómo en un momento determinado percibimos con claridad solo una pequeña parte de la escena. ¿Pero esto es algo nuevo? Antes de comenzar esta sección cualquiera de nosotros ya sabía que hay elementos de la escena que no podemos percibir en un determinado momento, por ejemplo, lo que hay detrás de nosotros. Aunque es interesante cómo nuestra percepción “real” u “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. Además, estas limitantes no implican que no podamos percibir bien los elementos específicos del mundo sobre los que prestamos atención, al igual que no tener ojos en la espalda no significa que no podamos darnos vuelta para ver lo que tenemos detrás. Y aunque los ejemplos discutidos son llamativos, es importante notar que el mundo real es muy diferente que las manipulaciones experimentales que muestran estos efectos (*e.g.*, en el mundo real los objetos no desaparecen repentinamente como en los experimentos de *change blindness*). A partir de esto la siguiente pregunta natural es: aquella parte de la escena sobre la que prestamos atención y que vemos con detenimiento, ¿es como 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 dos dimensiones (el plano de nuestra retina, al igual que las dos dimensiones de una fotografía), mientras que el mundo exterior tiene tres dimensiones espaciales. Nuestro sistema visual es muy bueno haciendo interpretaciones (subconscientes) de los estímulos visuales, dificultándonos apreciar las ambigüedades que tienen, pero estas se dan en los diversos pasos del procesamiento visual.

Uno de los primeros pasos 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. Pero esto no es fácil. Podemos pensar algunas reglas simples para lograrlo (*e.g.*, 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 cumple (*e.g.*, una superficie texturada con líneas marcadas, como una prenda de vestir). Esta etapa de procesamiento, que parece simple, en verdad implica una compleja interpretación de la estructura básica de la imagen y 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 (*e.g.*, a qué distancias se encuentran de nosotros, su orientación espacial, su tamaño, entre otras) y agruparlas en objetos. Continuando con el procesamiento construimos una escena en tres dimensiones de los objetos en nuestro entorno. Todos estos pasos involucran etapas de interpretación (*e.g.*, interpretar si dos “píxeles” son de la misma superficie o no) para llevarnos de la imagen bidimensional a un modelo tridimensional de la escena. Una imagen famosa que ilustra este proceso es la que se muestra en la Figura 7.5. Aunque es una imagen compuesta de manchas negras, nuestro cerebro logra generar un agrupamiento de los elementos, una composición de la estructura de las superficies, agruparlas en objetos y finalmente en una escena. Por su apariencia “desorganizada” este ejemplo es ilustrativo, ya que es difícil captar qué evidencia usa nuestro cerebro para interpretar ese caos de píxeles. Pero más allá

de los detalles del procesamiento, aquí nos importa que en el mismo puede haber más de una interpretación posible, lo que ilustran los siguientes ejemplos.

Figura 7.5



Interpretación de un conjunto de manchas dispersas. Esta figura muestra una clásica imagen de una escena que se pasó a blanco y negro, quedando con la apariencia de un conjunto de manchas negras dispersas. Pero a pesar de que la imagen es caótica, nuestro cerebro lleva a cabo una interpretación de este conjunto de manchas que le permiten ver la escena tridimensional de un dalmata caminando hacia un árbol.

La habitación de Ames es una ilusión clásica en la que generamos una interpretación errada de una escena. Este es un cuarto con una forma particular: no es rectangular, y tanto sus paredes como el techo y el piso están inclinados y no son paralelos. Para generar la ilusión, una de sus paredes tiene un agujero cuidadosamente ubicado a través del cual podemos mirar y percibir un cuarto rectangular normal (con paredes, techo y piso paralelos). Cuando miramos a través del agujero entonces nuestra percepción está errada, ya que creemos ver un cuarto normal cuando en verdad tenemos enfrente un cuarto distorsionado. Incluso, si hay gente en el cuarto, según su posición los percibiremos como gigantes o como pequeños. Lo importante a notar aquí es que el estímulo visual en este caso (y en todos los otros) es ambiguo. Es compatible tanto con un cuarto rectangular (como percibimos erróneamente), como con la escena real del cuarto irregular (ya que esta es en efecto 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, y la interpretación de la imagen implica elegir una de estas alternativas.

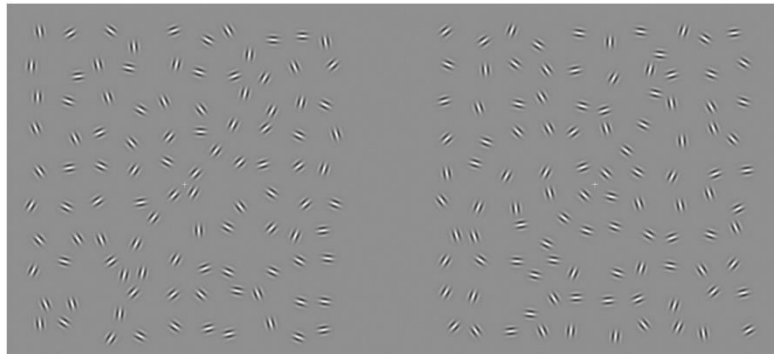
Otro ejemplo interesante, que no involucra la distribución espacial de la escena, es la ilusión de Adelson. En la misma se muestra la imagen de una superficie con un patrón como el de un tablero de ajedrez, con cuadrados blancos y negros. Arriba de la misma hay un objeto que proyecta una sombra sobre 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 de tal manera 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.

Estos ejemplos muestran que nuestro sistema visual lleva a cabo interpretaciones de los estímulos visuales para percibir la escena que nos rodea, pero que estas interpretaciones pueden equivocarse. ¿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 experimentos simplificados 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 cuál 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; Figura 7.6). 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 iluminada desde abajo, y viceversa (esto puede verse, por ejemplo, dando vuelta una foto de un cráter). Ante este tipo de ambigüedad nuestro sistema visual tenderá a elegir la escena en que la iluminación viene desde arriba, y luego interpretará si la superficie es cóncava o convexa según este criterio. 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 7.6



El agrupamiento de los elementos de la imagen es un proceso fundamental en la percepción visual. Este proceso se da según reglas complejas que aún no comprendemos completamente. En la imagen de la izquierda se muestra un ejemplo en que nuestro sistema visual agrupa los elementos colineales a pesar de encontrarse rodeados de otros elementos similares. En la imagen de la derecha tenemos los mismos elementos pero ortogonales a la dirección de alineamiento y nuestro sistema visual no los agrupa.

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, ¿por qué percibimos de forma errónea la habitación de Ames o la ilusión de Adelson? La respuesta es que son estímulos atípicos que violan la estructura normal del mundo. La imagen de la habitación 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 ocurre en el mundo real. Es mucho más probable que una imagen de ese tipo sea generada por un cuarto rectangular con sus paredes, techo y piso paralelos. De la misma forma, en la ilusión de Adelson la imagen es una foto en dos 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 identificar el color del cuadrado en sí). Aunque estas ilusiones muestran las interpretaciones del cerebro en funcionamiento, 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, con lo que percibir implica un proceso de interpretación. 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.

Entonces, ¿cuál es la respuesta a nuestra pregunta?

Vimos tres motivos neurobiológicos por los que el mundo podría no ser como lo vemos: 1) nuestro sistema visual no evolucionó para permitirnos ver al mundo como es, sino para permitirnos guiar acciones útiles; 2) representar al mundo tal 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 disponible parece decir que nuestra percepción es buena); 2) asignamos muy bien nuestros recursos de procesamiento y, aunque parte importante del mundo se 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 como lo vemos dependerá de la definición precisa de dicha 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) que 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 (*e.g.*, 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”.

LECTURAS RECOMENDADAS

- Dennett DC & Mittwoch U (1996): *Darwin's dangerous idea: Evolution and the meanings of life*. Annals of Human Genetics 60(3): 267-267.
- Lettvin JY, Maturana HR, McCulloch WS & Pitts WH (1959): *What the frog's eye tells the frog's brain*. Proceedings of the IRE 47(11): 1940-1951.
- Shapiro AG & Todorović D (eds.) (2017): *The Oxford compendium of visual illusions*. Oxford University Press, New York.

¿Es el mundo como lo vemos?
